行政院國家科學委員會專題研究計畫 成果報告

影響不同性別學生數學學習相關因素之發展趨勢,與可能模式之建立與驗證(第 3 年)

研究成果報告(完整版)

計 畫 類 別 ： 個別型
計 畫 編 號 ： NSC 96-2522-S-003-017-MY3
執 行 期 間 ： 98 年 08 月 01 日至 100 年 07 月 31 日
執 行 單 位 ： 國立臺灣師範大學特殊教育學系（所）

計 畫 主 持 人 ： 陳心怡
共 同 主 持 人 ： 陳美芳
計畫參與人員 ： 大專生-兼任助理人員：劉昀青
大專生-兼任助理人員：孫采微
大專生-兼任助理人員：林冠坊
大專生-兼任助理人員：柯譯淇

報 告 附 件 ： 出席國際會議研究心得報告及發表論文

處 理 方 式 ： 本計畫涉及專利或其他智慧財產權，2 年後可公開查詢

中 華 民 國 100 年 10 月 28 日

Models for predicting adolescents' math achievement and math career intention across ages 9 to 15: Developmental trends of gender similarities and gender differences

Abstract

The purpose of this study was to investigate the developmental trends of gender similarities and differences on factors which affect boys and girls' math-achievement and math-career intention. Based on comprehensive data from a total 3,157 adolescents ranging in age from 9 to 15 years old, latent factor SEM models for students with different age level and gender were established. The major findings were: 1) Gender has small to moderate effect on math performance and career intention. All effects are indirectly mediated by other factors. 2) Ability factors are important predictive factors for math performance, while math interest, math self-efficacy, perceived peer support, math anxiety, outcome expectancy and personal style better predict math career intention. Math performance and math career intention are distinct constructs, with only low to median correlation. 3) The way boys and girls differ varies as age grows. Generally, a stable link between constructs for all samples was from math anxiety, to math self-efficacy, math interest, and math career intention. Besides, stereotype, math anxiety, and math self-efficacy were found to show comparatively more strong impacts for girls. 4) Family support, organized style, and perceived peer support are constructs which show many links to other learning-related variables. 5) At some age point from Grade 4 to grade 7, girls' math-anxiety and math-gender stereotype grows rapidly. This is the critical period which deserves educators' very attention. 6) As age grows older, the relative importance of family support and interest decrease, while the relative importance of self-efficacy and ability increase. Applications and future directions based on these findings were discussed.

Keywords: Developmental approach, Gender similarities, gender differences, Math achievement, Math career intention, SEM.

台灣9-15歲學生數學成就與數學生涯選擇意願之預測模式建立:
性別差異 與 發展趨勢分析

摘要

本研究的目的在分析影響男女學生「數學成就」與「數學生涯選擇意願」之多元變因, 了解變因間彼此影響路徑關係, 並比較在小學, 國中,高中三個不同發展階段內之趨勢異同。 研究團隊在經過嚴謹且大樣本的預試選題過程後, 經由多次實際施測共蒐集台北地區 10 所學校 3,157 位 9-15 歲學生的多元資料。資料內容涵蓋每位學生的推理能力,空間關係能力, 數學成就表現, 及與學習意動面向相關問卷。研究者將觀察變項經因素分析組成 22 個潛在因素(latent factor)以排除誤差影響, 並經由潛在變項結構方程模式(latent factor SEM)進行模式建立與驗證。變項內容包含認知, 人格,動機,興趣, 環境,與數學-性別刻板印象等完整向度。結果得以建立國小,國中, 高中三個階段男女學生共六個數學學習預測模式。研究主要發現為: 1) 性別對於「數學成就」與「數學生涯選擇意願」之影響有小到中度效果值。且所有影響都是間接影響, 非直接影響! 此表示中介變因的分析實為重要。2) 認知能力是最能預測「數學成的變因, 且其影響多為直接影響。 但興趣, 自我效能, 同儕支持, 焦慮程度, 結果期望,與人格傾向等向度才是能有效預測「數學生涯選擇意願」的主要變項。「數學成就」與「數學生涯選擇意願」兩者具低至中度相關, 二者是不同建構, 需被分開探討. 3) 男女學生所展現的性別差異隨著年齡上升而有不同。所有年段男女學生都展現的穩定顯著變項影響路徑是「數學焦慮」會影響「數學自我效能」, 進而影響「數學興趣」, 之後便影響「數學生涯選擇意願」。 此外, 數學-性別刻板印象 與數學自我效能對女學生的影響相對較大。 4) 家庭對於學習的支持, 學生本

身的組織性習性(personal style),以及知覺同儕對數學學習的支持等因素是所有變項中對有最廣泛影響的幾個因素. 5）由小學中高年級到中學一年級之間,是女學生數學焦慮， 與數學-性別刻板印象顯著上升的關鍵時刻,特別值得教育工作者留意。6）就「數學生涯選擇意願」而言，隨著年齡上升， 家庭支持與數學興趣的重要性相對下降， 但認知能力與自我效能的之重要性則相對上升。文中對於研究發現有細步討論,並對未來研究提出建議。

關鍵字: 數學生涯選擇意願，數學成就，性別差異,發展趨勢，潛在變項結構方程模式.

Introduction

Math and science is, without doubt, fundamental and crucial. However, one consistent finding has been the gap in standardized tests of mathematics favoring males (Ackerman, 2006; Halpern, 2000; Halpern, Wai, & Saw, 2005; Maccoby & Jacklin, 1974). Even if the gender gap in mathematical and scientific performances is closing, there still no parity in representation of males and females in the science-related professional fields (Spelke, 2005; Wickware, 1997).

Various biological and social variables vary as a function of gender. Some of these differences may have substantial importance and consequences (Davis & Shackelford, 2006; Zuriff, 2006). Lippa (2006) suggested that, a balanced perspective recognizing both gender differences and gender similarities, or so called 'gender reality' better be hold. We find this point by Halpern (2000, p.8) most convincing "differences are not deficiencies, and it is only through careful study of differences that similarities can be revealed".

Based on the meta analyses by Hyde(2005), males excel females on overall math performance(effect size d=+.16) 、 math problem solving (d=+.08) 、 math self confidence(d=+.16); On the contrary, females excel males on computation(d=-.14) 、 number(d=-.10), and show more math anxiety (d=-.15). There seemed no gender differences on math concepts (d=-.03). These results show that males perform better on math works related complicated higher-level information processing, while females perform better on math which need calculations. Besides, researchers also suggested a need to study this gender issue in a developmental view (Hyde, Fennema, & Lamon, 1990).

Besides gender, there are multiple factors which affect learning, such as cognitive abilities, motivation, interest, attitude, etc. Several researchers suggested that when considering gender-math issue, all the learning related factors better be considered into the same picture. For example, Bryne's (2003,2005) 'Three Conditions Model (3C model)' suggested that learning opportunities, motivation, and aptitude are the three major factors for explaining math performance. Other factors such as gene, learning experiences, and family/school environments are variables which affect those three major factors. This 3C model explains roughly 40-50% variance on math performance. Matthews, Zeigner, & Roberts (2006) also raised another hypothesized model for learning performance. Variables from multi-domains, such as intelligence, aptitude, domain knowledge, personality, motivation (self-efficacy), efforts, and anxiety, were all considered as crucial ones.

Math learning is known to associate with all these abovementioned variables. Math performance is reported to related to intelligence (Ackerman, Bowen, Beier, & Kanfer,2001; Glutting, Watkins, Konold, & McDermott, 2006), working memory (Keeler & Swanson,2001; Swanson & Beebe-Frankenberger,2004), self-concept(Marsh & Hau,2004), self-efficacy (Kenney-Benson, Pomerantz, & Ryan, 2006), math anxiety (Chipman, Krantz & Silver,1992, 1995), math interest (Ackerman, Bowen, Beier,& Kanfer, 2001, 2002), and parents support (Brown & Josephs, 1999). Since these factors all related to math learning, they could be mediating variables which explain the effect of gender on math learning. For example, researchers suggested that visual-spatial and mental rotation ability is strongly associated with gender difference on math (Geary, Saults, Liu, & Hoard, 2000; Halpern, Wai,& Saw, 2005; Hedges & Nowell, 1995; Nuttall, Casey,& Pezaris, 2005). Gender differences also have been found in reasoning (Lynn & Irwing, 2004, 2005), personality (Feingold, 1994; Pinquart & Sörensen, 2001), math anxiety (Hyde, Fennema, Ryan, Frost, & Hopp, 1990), self-efficacy(Ewers & Wood, 1993; Pajares & Miller, 1994), math interest/math career intention(Wigfield, Battle, Keller, & Eccles, 2002), and math experiences provided by parents (Ruble & Martin, 1998).As suspected, all these discrepancies may somehow work together for accounting the gender difference on learning outcome (Ackerman, Bowen, Beier, & Kanfer , 2001). It is thus important to differentiate all these paths while studying gender reality on math (Ackerman & Lohman, 2006;

Chipman,2005).

In our recent study (Chen, Chen, Chang, & Lee, 2009), we investigated developmental trends in gender reality for the school-age children in Taiwan. Data sets for 11 cognitive and affective psychological tests were analyzed (altogether, 17,453 males and 16,526 females), majority of them are large, representative, and normative data. Our results supported the importance of viewing gender reality from a developmental perspective. Most importantly, gender differences in affective attributes such as personality, interest, and learning styles were fairly stable across age levels (d = 0.30 to 0.80). Cognitive advantages for each gender, however, varied with developmental phase. We were astonished at how distinct gender affective attributes were extraordinarily constant across age levels. Styles of personality, learning styles, emotions, and interests expressed in early elementary school seemed to remain unchanged as the children developed. Generally, girls showed slightly stronger levels of depression and had more organized and feeling-oriented learning styles. Boys expressed more imaginative, flexible, and thinking-oriented learning styles and showed more rule-violating behaviors. Interests of each gender were also quite different. Boys show stronger preferences for mechanical and scientific activities. Girls find people-oriented activities more attractive, such as teaching, persuasion, and social services. In the gender-math literature, these variables have seldom been investigated in the hypothesized model. However, we suspected that they could be important mediating factors.

In fact, we should not under-estimate possible cumulative effects of all these factors on math gender-differences. For example, given that the gender gap in mathematical and scientific performances is closing, why is there still no parity in representation of males and females in the science-related professional fields (Spelke, 2005; Wickware, 1997)? As findings demonstrated, on the average, males tend to develop to be more thinking–oriented and more flexible. They had better general knowledge, were more emotional stable, were better at reasoning and less interested in spending time teach and persuade others. Gridley (2006) pointed out that ability cannot explain everything. Thinking styles, such as thinking-feeling orientation, help an individual with career selection. One's preference and orientation toward people or thing may play a crucial role in the kind of career that one become interested in. As Feist (2006, p.163) contended, "Imaging a scientist without a unique style of behavior and thinking is nearly impossible. Scientific interest and achievement have fascinating and complex developmental paths and are more likely to come from people with particular kinds of personalities and traits than with other kinds of personalities". Webb, Lubinski, & Benbow (2002) also suggested the effect of individual differences on influencing human decisions cannot be ignored. Equal gender representation across all educational-vocational domains may conflict with what might be happening naturally. Thus, "equal male-female representation across disciplines may not be as simple to accomplish as many policy discussions imply (Webb et al. 2002, p.785)".

As Halpern, Benbow, Geary, Gur, Hyde, and Gernsbacher (2007, p.41) wonderfully concluded, "There is no single factor by itself that has been shown to determine sex differences in science and math. Early experience, biological constraints, educational policy, and cultural context each have effects, and these effects add and interact in complex and sometimes unpredictable ways". Since recent research consistently reported that the gender gap in mathematical and scientific performances is closing, however there still no parity in representation of males and females in the science-related professional fields (Spelke, 2005; Wickware, 1997). Considering all the above findings, we carefully selected variables in this current study for better understand the underlying mechanism for modeling gender reality on math performance and math career intention. Especially, affective factors which show large and constant gender differences were jointly valued and investigated.

Methods

Participants

We analyzed data from three samples: students from elementary school (4th graders), junior-high school (7th graders), and senior-high school (10th graders).

The elementary sample consists of 860 4th graders (9 years ole). They were from four schools in both Taipei city and New Taipei city. A total of 818 children (429 boys and 389 girls) were later considered as being valid cases for formal analyses because of showing acceptable percentages of missing responses. The average verbal intelligence (VIQ) of the sample was 106.2, with a SD of 14.0. The mean performance intelligence (PIQ) of the sample was 101.6, with a SD of 14.0.

The Junior-high school sample consists of 1,145 7th graders (12 years ole). They were from three schools in both Taipei city and New Taipei city. A total of 1,102 children (568 boys and 534 girls) were later considered as being valid cases for formal analyses because of showing acceptable percentages of missing responses. The average verbal intelligence (VIQ) of the sample was 102.6, with a SD of 15.5. The mean performance intelligence (PIQ) of the sample was 102.1, with a SD of 16.5.

The Senior-high school sample consists of 1,259 10th graders (15 years ole). They were from three schools in both Taipei city and New Taipei city. A total of 1,237 children (717 boys and 520 girls) were later considered as being valid cases for formal analyses because of showing acceptable percentages of missing responses. The average verbal intelligence (VIQ) of the sample was 107.20, with a SD of 12.3. The mean performance intelligence (PIQ) of the sample was 107.50, with a SD of 12.45. Obviously, compared to the national norm, high school students are group with higher abilities.

Instrumentation

For each participant, data on four domains were collected via the instruments described as follows:

(a) Reasoning ability:

The Otis-Lennon School Ability Test (OLSAT) (Otis & Lennon, 2006, 2008) was used to measure both the verbal and nonverbal reasoning abilities (or also called VIQ and PIQ). Each domain contains 30 items, and altogether takes 40, 45, and 50 minutes to complete for 4thm 7^{th} and 10^{th} graders. The standardized and representative Taiwan norm was recently developed, and it was demonstrated to have good reliability (internal consistency is around 0.90 for all ages) and validity.

(b) Spatial ability:

The spatial mental rotation subtest in Differential Aptitude Tests-V (DAT-V) (Bennett, Seashore, & Wesman, 1999) was used. This subtest has 50 items, and the time limit was 15 minutes, as specified in the standardized manual. The Taiwan version reported both good reliability (with internal consistency above 0.90 for all ages) and validity. A total of 240 students were pilot tested for ensuring the utility of this test in this current study.

(c) Math achievement:

The various forms of math achievement test (one form for each grade) were developed by the authors and an expert team including professors in the area of Mathematic education, and experienced math teachers. The process of test development followed standard psychometric procedure: an expert team of six professionals was first grouped. The structure of TIMS 2007 was reviewed and followed, and the construct of each test form was decided to be specified by two major domains: cognitive domain and content domains. Cognitive domain includes 3 categories: knowing, Applying, and Reasoning; while the content domain includes 5 categories: Number, Measurement, Geometry, Algebra, Data and chance. Based on the official Mathematics curriculum outline announced by Ministry of Education, the expert team decided the percentages of each categories (see appendix 1 for detailed percentage information for each form), which were then served as the guide for item writing and test development.

Pilot testing was conducted with a sample of 1,306 children from age 8 to 16. Problematic items were either deleted or revised. Each form contains 20 to 30 items with multiple types (multiple choice, fill in the blank, open-ended questions, etc). Each form take about 40, 45, and 50 minutes to complete for 4th, 7th , and 10th graders. The reliabilities of final forms are around 0.62 to 0.86 (Md = .83).

(d) Other math-learning related personal factors:

A self-reported questionnaire (with likert-type, five-point items) was developed by authors. Preliminary items were written by authors based on important constructs from a comprehensive literature review. We reviewed literatures in fields of gender research and general learning related theories. Variables from several theories were carefully considered and selected. The main referenced theories were: Theory for Educational Productivity (Walberg, 1984, 1988); Social Cognitive Career Theory (SCCT) (Lent, Brown, Hackett, 1993, 1994, 2000; Lent 2003); and Self Determination Theory (SDT) (Deci &Ryan, 1985, 2000, 2008).

At the preliminary phase, the 1st draft of this questionnaire included 182 items which constitute 26 factors (7 items for each factor). After item analyses based on a pilot sample of 1,245 children, only 100 items which contributing to 25 factors remained (4 items for each factor). The median factor reliability was .75 based on pilot sample.

Briefly, questionnaire data covers the following main categories: (1) personal affective characteristics (thinking vs feeling oriented learning styles,; organized vs flexible oriented learning styles, negative emotions, positive self concept, people-oriented interest , and tendency to obey rules); (2) family background and perceived parents' factors (expectation, involvement, autonomy supportive, and math-learning supportive); (3) perceived math teachers' factors (autonomy supportive, expectation and teaching involvement); (4) perceived peer' factors (math-learning supportive); (5) perceived math-related environmental factors (relatedness, others' math-gender stereotype); (6) previous math-learning experience; (7) personal motivation factors (self efficacy, outcome expectancy, intrinsic and extrinsic motivations, and efforts); and (8) personal math attitude (self math-gender stereotype, math anxiety, math interest, and intention to choose math as career options).

Later on, we conducted more advanced factor analyses based on formal sample, and revised the composite of studied factors. The final selected factors from this questionnaire were: 1) Parent education level (2 items); 2) Family support /involvement (5 items); 3) feeling-oriented style (5 items); 4) organized-oriented style (5 items); 5) negative style (5 items); 6) human-oriented style (5 items); 7) perceived others' math-gender stereotype (4 items); 8) self math-gender stereotype (4 items); 9) math-anxiety (4 items); 10) parents' math support (3 items); 11) peers' math support (4 items); 12) math teachers' teaching quality/support (8 items); 12) math self-efficacy(4 items); 14) math outcome-expectancy (4 items); 15) math interest (5 items); 16) study efforts made for math (4 items); and 17) math career intention(4 items)

Hypothesized starting model

Based on a through literature, the baseline starting model, which predicts the math performance and math career intention, was first hypothesized. A total of 22 factors were selected for modeling (gender, VIQ, PIQ, spatial ability, math performance, and the other 17 factors from questionnaire as mentioned in the above session).

Figure 1 shows the hypothesized starting model. In this model, gender was proposed to have direct effect on ability, personality, and perceived environment support. Relevant background characteristics were controlled, including parent education level and family involvement. It was assumed that gender differences on math performance and math career intention were indirectly mediated by intervening factors such as abilities, personal style, perceived environment support, stereotype, anxiety, self-efficacy, outcome expectancy, interest, and effort. The unique characteristic of this model was that a comprehensive background factors (abilities, personality and styles, family background) were jointly considered, which will make us having a better control and better understanding on the relationships between other factors(Keith, 2006). Detailed variable quality will be reported in the result session.

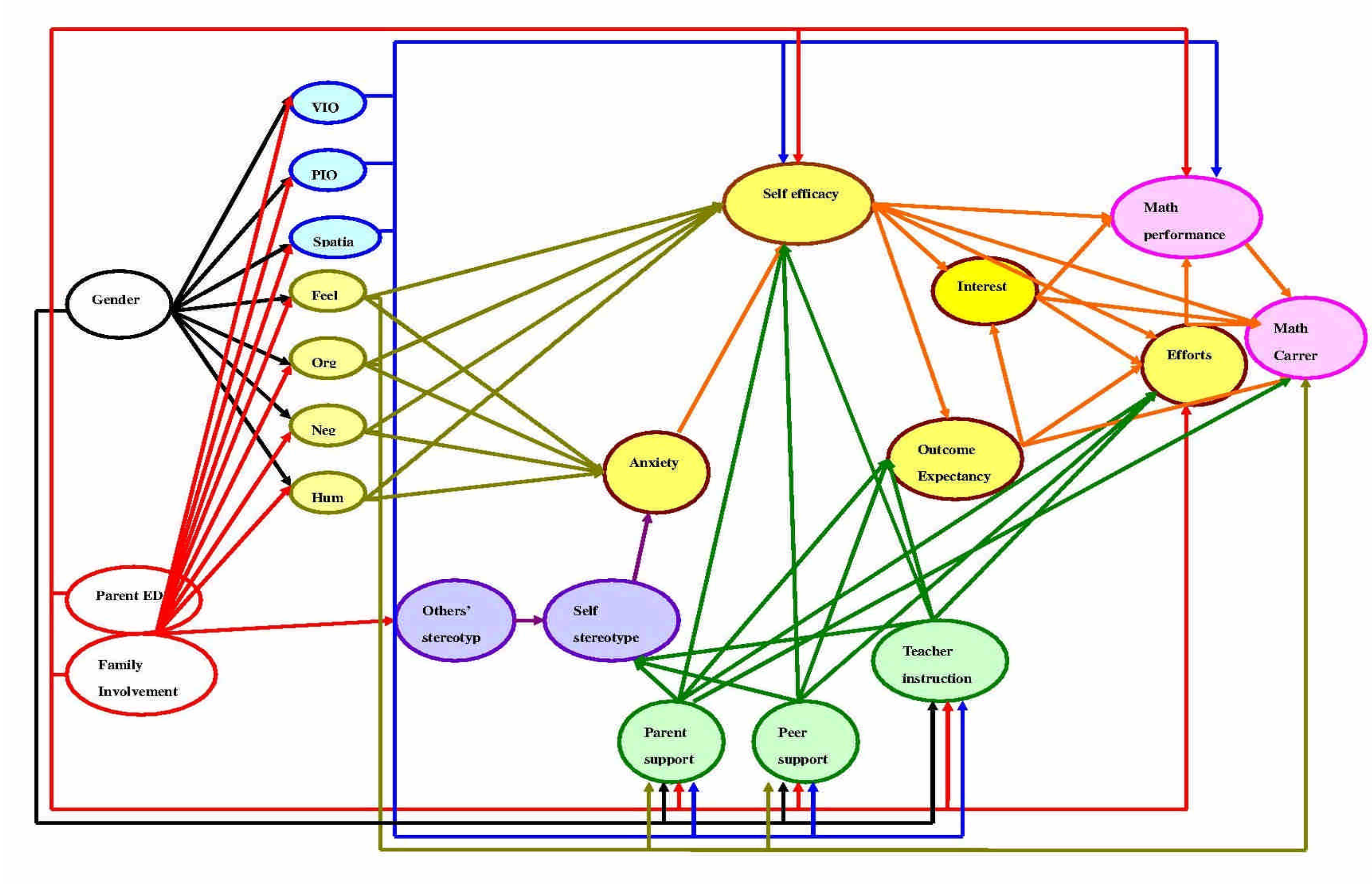


Figure 1. The hypothesized starting model

Analysis

Latent variable structural equation modeling (SEM) was used to determine the magnitude of the influence of relevant variables on Math performance and career intention.

Based on this same specified starting model as depicted in figure 1, we analyzed the data from grade 4, grade 7, and grade10 separately. Thus a total of three sets of analyses were processed. For each set of analyses, a calibration-validation approach was used where two-third of each sample was randomly selected as the calibration sample to test hypotheses and modifications (modifications were guided by both theoretical meaningfulness and LISREL MI index), while remaining third was used to cross-validate the results of calibration analyses. Once a best-fitting solution for each data set (grade 4, 7, or 10) was calibrated and validated, final parameters were retested using the entire sample. For model clarity, only paths showing at least small effects ($\beta \geqq .05$) were kept for later separated analyses for each gender.

All the SEM runs were conducted based on the analysis of covariance structure models using LISREL 8.8 (Jöreskog & Sörbom, 2006). The scale of latent factors were defined by fixing a factor loading each to one. Multiple indices of model fit (Bentler & Bonett, 1980; Hu & Bentler, 1998, 1999; Kline,2005; Marsh, Balla, & McDonald, 1988) helped us evaluate and compare the various models in this study. Single models were evaluated using comparative fit index (CFI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). An RMSEA less than .05 corresponded to a good fit and with .08 considered an acceptable fit (McDonald & Ho, 2002). For completeness, we included the 90% confidence interval for RMSEA. The SRMR values less than .08 were considered acceptable (Hu & Bentler, 1999). A value of 0.90 served as the rule-of thumb lower limit of acceptable fit for all indices ranging from zero to 1, with 1 indicating a perfect fit (Hoyle & Panter, 1995; Kline, 2005). Change in chi-squared ($\Delta\chi^2$) evaluated competing, nested models (Bentler & Bonett, 1980). Akaike information criterion (AIC) and sample size adjusted Bayesian Information Criterion (aBIC) helped with comparisons of non-nested models (Kaplan, 2000; Loehlin, 2004), with smaller values indicating a better fit. Comparatively, aBIC has a greater reward for parsimony than does the AIC. If inadequate fit was detected, fit in the model was improved by including additional parameters identified by the modification index (MI) provided by LISREL. Re-parameterization will be examined carefully for meaningfulness.

For explanation, suggestions by Keith (2006) were used for quantifying the magnitude of effects. For studies in the learning field, "β's above .05 are considered small but meaningful; those above .10 are considered moderate, and those above .25 are considered large." (Keith, 2006, p.62)

Results

1. The measurement model

The measurement model was specified and tested before interpreting the relations between latent factors based on structure-estimation. Table 1 contains the reliabilities of all 21 studied factors (except variable gender), and factor loadings of each corresponding manifest indicators. Altogether, if “gender” was included, a total of 80 manifest indicators constitute the 22 studied latent factors. In this analysis, majority of the latent factors were identified by 4-5 manifest items (the range is 3 to 8 items). The only exceptions were latent factors identified by single indicators (total score of corresponding test) such as VIQ, PIQ, spatial ability, and math performance.

As shown in the table, the average reliability for grade 4, 7, and 10 were .76, .82, and .80 accordingly. The median reliability was .78, .81, and .78 accordingly. The factor loadings (β) are standardized maximum likelihood coefficients. All factor loadings were significant! The findings of moderate- to high- factor loadings of the indicators for most factors, suggested that these selected items and tests were effective instruments in defining the latent variables. Latent factors in the starting model were measured in a valid manner. The acceptable level of measurement validity was achieved.

Table 1. Reliability of latent constructs and corresponding factor loadings of manifest indications

Latent Construct	Manifest indicators (item contents)	Grade 4		Grade 7		Grade 10	
		α	β	α	β	α	β
1. Parent educational level	fathers' educational level	.80	.71	.79	.81	.77	.94
	mothers' educational level	---	1.05	---	.87	---	.68
2. Family support	parents ask about school learning often	.72	.49	.82	.51	.80	.40
	parents discuss with me about life problems often	---	.51	---	.56	---	.66
	parents understand me	---	.49	---	.75	---	.80
	parents try to understand my thoughts	---	.54	---	.75	---	.72
	I can trust my parents	---	.57	---	.75	---	.75
3. Verbal IQ	Score on Standardized OLSAT verbal scale	.76	.87	.79	.89	.71	.85
4. Performance IQ	Score on Standardized OLSAT performance scale	.80	.89	.86	.93	.81	.90
5. Spatial ability	Score on Standardized Differential Aptitude Test: Spatial subtest	.87	.93	.89	.94	.92	.96
6. feeling-oriented style	I feel unhappy when seeing others feel unhappy	.54	.36	.69	.50	.71	.58
	I feel being influenced when watching sad Movie senario	---	.38	---	.48	---	.56
	I feel sad too when my team members are encountering some bad news and feel said	---	.41	---	.67	---	.69
	Compared to my friends, I am a person who is more easily touched	---	.54	---	.68	---	.70
	I used to play a role of taking friends complaints and encourage others	---	.48	---	.47	---	.43
7. Organized style	I prefer my study desk neat and well-organized	.57	.41	.69	.57	.62	.47
	I prefer work first, and have fun later	---	.44	---	.62	---	.63
	When lots works are to be done, I will think ahead the working sequence, and then proceed	---	.56	---	.53	---	.54
	I am a person who obey regulations and rules	---	.57	---	.55	---	.49
	I do not do things which is violating the rule	---	.27	---	.47	---	.27
8. Negative style	I feel nervous often	.61	.64	.69	.62	.72	.57
	I feel worry often	---	.65	---	.82	---	.78

	I feel sad often	---	.75	---	.74	---	.70
	I am not satisfied with myself	---	.12	---	.27	---	.42
	I can not do things well often	---	.34	---	.32	---	.44
9. Human-oriented style	Compared to my friends, I am more optimistic	.79	.41	.79	.42	.78	.36
	Others like to be with me	---	.56	---	.49	---	.43
	I like to be with others	---	.77	---	.83	---	.82
	I feel happy when helping others	---	.74	---	.72	---	.66
	It is interesting to have interaction with others	---	.81	---	.85	---	.90
10. Perceived others' math-gender stereotype	My parents think that male, but not female, should choose science/engineering related majors	.77	.55	.78	.62	.75	.68
	Teachers make me feel that math is more important for male, and it does not matter for girls performing poorly on math	---	.77	---	.69	---	.72
	Most of my friends feel that girls are good at literacy, and math is not the strong subject for female	---	.60	---	.59	---	.45
	My parents feel that learning math well is useful for male, and less useful for female	---	.72	---	.77	---	.78
11.Self math-gender stereotype	male can solve math question faster and more accurate	.83	.72	.90	.77	.87	.76
	males are bone to have better math ability	---	.84	---	.89	---	.82
	male should choose majors which are math or science/engineering related	---	.61	---	.84	---	.77
	It is hard for female to outperform male on math	---	.76	---	.83	---	.80
12. Math Anxiety	I feel nervous(worried, scaled) when thinking about math class	.83	.71	.85	.77	.82	.74
	I will not able to comprehend the instruction from math teacher because of nervous and worry	---	.66	---	.74	---	.63
	I feel nervous for no reason when facing a lot of number, figures, and tables	---	.74	---	.79	---	.76
	when math teacher is asking questions, I feel nervous and uncomfortable	---	.78	---	.78	---	.80
13. perceived Math support	parents care my math exam scores	.63	.64	.72	.70	.72	.67

from parents							
	parents value my math learning	---	.77	---	.93	---	1.03
	parents will find ways to teach me math homework (either themselves/ home tutor/ or afterschool class)	---	.26	---	.39	---	.39
14. perceived Math support from peers	I will discuss math questions with my friends	.66	.67	.65	.55	.69	.72
	I discuss math questions with friends after class often	---	.19	---	.37	---	.56
	Friends and I will encourage each other to learn math well	---	.62	---	.65	---	.68
	I feel respected and supported by friends in math class	---	.63	---	.52	---	.43
15.perceived teaching quality of math teacher	math teacher will try to understand our thoughts	.82	.66	.86	.70	.88	.79
	I feel we are valued and respected by math teacher	---	.61	---	.73	---	.84
	math teacher knows us	---	.40	---	.62	---	.74
	math teacher encourage us to express our opinions	---	.57	---	.60	---	.65
	in math class, teacher cares our comprehension	---	.36	---	.52	---	.68
	math teacher explain where we got wrong carefully after exam	---	.63	---	.55	---	.51
	math teacher encourage us to getting better each time	---	.71	---	.61	---	.58
	I feel math teacher makes great efforts to teach us math	---	.72	---	.72	---	.70
16. Math self-efficacy	I am confident in my math learning ability	.78	.68	.88	.75	.89	.78
	Math is the subject which I am good at	---	.61	---	.74	---	.86
	Learning math is easy for me	---	.63	---	.81	---	.76
	I feel my math ability is good	---	.65	---	.77	---	.80
17.Math outcome expectancy	I will have better ability for Colleague, if I studied math well	.73	.83	.81	.77	.79	.63
	I may able to get a higher-paid job, if I studied math well	---	.77	---	.83	---	.74
	learning math well, is importantly related to my future life	---	.35	---	.53	---	.62
	I can be in the job I prefer, if I studied math well	---	.61	---	.71	---	.76
18. Math interest	I want to learn math well because it is interesting	.85	.67	.91	.77	.92	.80
	I want to learn math well because I like to solve math problems	---	.70	---	.77	---	.76
	I like math	---	.77	---	.82	---	.86
	I feel learning math is enjoyable	---	.73	---	.87	---	.86

	I am interested in learning math	---	.73	---	.78	---	.83
19. Math efforts	I study the math afterschool often	.69	.31	.74	.44	.70	.52
	If my math score was not good, I work harder for it	---	.68	---	.76	---	.64
	I try hard to do math homework	---	.70	---	.72	---	.67
	for math formula which are harder, I will try harder to memorize them	---	.62	---	.62	---	.52
20. Math performance	test score of the math achievement test	.83	.94	.89	.95	.72	.85
21. Math career intention	I hope to choose more math-related courses, if possible	.80	.72	.86	.72	.88	.79
	I hope to encounter math again after graduation	---	.67	---	.86	---	.83
	I hope to study math-related majors	---	.62	---	.83	---	.82
	I hope I can be in a math-related job in the future	---	.76	---	.71	---	.71
Reliability mean (Fisher r to z)		**.76**		**.82**		**.80**	
median		**.78**		**.81**		**.78**	

2. The structural model

(2.1) Elementary sample (Grade 4)

All examinations based on the elementary sample (grade 4) are shown in Table 2a In the calibration phase, goodness-of-fit indexes reported for the initial starting model (model 1) were within the acceptable range (CFI=.94, RMSEA=.051, SRMR=.074), showing the literature-driven, hypothesized starting model (which was shown in figure 1) fit data well. In the following modification process, reasonable parameters which need to be modified were checked one at a time, a total of thirty-six originally unspecified parameters were found to yield statistically significant improvement in model fit. The goodness-of-fit indexes reported for this modified model (model 1b) were improved (CFI=.96, RMSEA=.043, SRMR=.067).

Validation analyses tested this modified structure with a different data set (model 2). Results showed that the modified structure had an acceptable fit to the 2nd set of data (CFI=.93, RMSEA=.043, SRMR=.080). We thus test this validated structure by the entire elementary sample (N=818) in model 3, the results showed a good fit (CFI=.96, RMSEA=.041, SRMR=.064).

Because the validated-model is complicated with 22 latent factors and many paths, for clarification purpose, we decided to only keep the paths which shown at least a small effect ($\beta \geqq .05$). A total of 29 small or non-significant paths were removed, which released more degree of freedom, improved the parsimony, and yield a better fitting model (model 4) (CFI=.96, RMSEA=.040, SRMR=.064). This model was served as the main base for 4th graders, and the figure of this model was shown in figure 2a.

We further tested this model separately for each gender (the variable gender was removed from the model). Generally, the final model fit both gender well. With only a few modifications, the final modified model for boys' group (model 5, N=429) showed a good fit (CFI=.94, RMSEA=.046, SRMR=.074). Similarly, the final model for girls' group (model 6, N=389) also showed a good fit (CFI=.95, RMSEA=.040, SRMR=.069). Results seemed to reveal that the general leaning model for each gender is similar. However, it would be interesting to explore whether these variables have the same effect on learning for different genders. These questions will be discussed in the following sessions.

(2.2) Junior high school sample (Grade 7)

All examinations based on the junior high school sample (grade 7) are shown in Table 2b. In the calibration phase, goodness-of-fit indexes reported for the initial starting model (model 1) were within the acceptable range (CFI=.94, RMSEA=.048, SRMR=.070), showing that the literature-driven starting model (which was shown in figure 1) fit data well. In the following modification process, reasonable parameters which need to be modified were checked one at a time, a total of 26 originally unspecified parameters were found to yield statistically significant improvement in model fit. The goodness-of-fit indexes reported for this modified model (model 1b) were improved (CFI=.96, RMSEA=.041, SRMR=.060).

Validation analyses tested this modified structure with a different data set (model 2). Results showed that the modified structure had an acceptable fit to the 2nd set of data (CFI=.95, RMSEA=.042, SRMR=.070). We thus test this validated structure by the entire grade 7 sample (N=1,102) in model 3, the results showed a good fit (CFI=.96, RMSEA=.039, SRMR=.057).

Because the validated-model is complicated with 22 latent factors and many paths, for clarification purpose, we decided to only keep the paths which shown at least a small effect ($\beta \geqq .05$). A total of 33 small or non-significant paths were removed, which released more degree of freedom, improved the parsimony, and yield a better fitting model (model 4) (CFI=.96, RMSEA=.039, SRMR=.058). This model was served as the main results for 7th graders, and the figure for this model was shown in figure 2b.

We further tested this model separately for each gender (the variable gender was removed from the model). The final model fit both gender well. With a few modifications, the final model for boys' group (model 5, N=568) showed a good fit (CFI=.96, RMSEA=.042, SRMR=.067). Similarly, the final model for girls' group (model 6, N=534) also showed a good fit (CFI=.96, RMSEA=.038, SRMR=.063). For this group, more modifications (identifying extra significant paths) seemed necessary for boys' group than for girls' group. Thus results seemed to reveal that the general background leaning model for each gender is similar. However, for students at this age, gender differences seemed become more salient somewhat. In the following sections, we will explore whether these variables have the same effect on learning for different genders.

(2.3) Senior high school sample (Grade 10)

All examinations based on the senior high school sample (grade 10) are shown in Table 2c In the calibration phase, goodness-of-fit indexes reported for the initial starting model (model 1) were within the acceptable range (CFI=.93, RMSEA=.050, SRMR=.072), showing the hypothesized starting model (which was shown in figure 1) fit data well. In the following modification process, reasonable parameters which need to be modified were checked one at a time, a total of 40 originally unspecified parameters were found to yield statistically significant improvement in model fit. The goodness-of-fit indexes reported for this modified model (model 1b) were improved (CFI=.95, RMSEA=.040, SRMR=.055).

Validation analyses tested this modified structure with a different data set (model 2). Results showed that the modified structure had an acceptable fit to the 2nd set of data (CFI=.94, RMSEA=.040, SRMR=.066). We thus test this validated structure by the entire grade 10 sample (N=1,237) in model 3, the results showed a good fit (CFI=.96, RMSEA=.039, SRMR=.053).

Because the validated-model is complicated with 22 latent factors and many paths, for clarification purpose, we decided to only keep the paths which shown at least a small effect ($\beta \geqq .05$). A total of 27 small and non-significant paths were removed, which released more degree of freedom, improved the parsimony, and yield a similar good fitting model (model 4) (CFI=.96, RMSEA=.039, SRMR=.053). This model was served as the main results for 10th graders, and was shown in figure 2c.

We further tested this model separately for each gender (the variable gender was removed from the model). The final model fit both gender well. With only a few modifications, the final model for boys' group (model 5, N=717) showed a good fit (CFI=.95, RMSEA=.041, SRMR=.057). Similarly, the final model for girls' group (model 6, N=520) also showed a good fit (CFI=.95, RMSEA=.037, SRMR=.059). Results supported that the general leaning model for each gender at this age level is similar. We explore the effect of them on learning for different genders in the next sessions.

In this research, our models not only explain data well. Because each factor is carefully selected and pilot tested, we were able to explain large portion of the variances of important dependent variables. For example, our models explain roughly over 90% of the math performance variance for 4th and 7th graders; while it was about 43-61% for the 10th graders (noticing that our 10th graders are a group of higher ability students). Besides, current models explain over 90% of the math career intention variance for 4th graders (it was due to the high correlation between math interest and math career intention), 81-84% of the total career intention variance for the 7th graders, and 88-92% of the total career-intention variance for the 10th graders.

Table 2a. Hypotheses testing for the Elementary sample (Grade 4):

model	χ^2	df	χ^2/df	CFI	RMSEA	RMSEA 90%CI	SRMR	AIC	aBIC
1. Calibration sample (N=545)- starting	7275.02	2984	2.44	.94	.051	.050-.053	.074	7787.02	
1b. Calibration sample (N=545)-modified	5879.07	2948	1.99	.96	.043	.041-.044	.067	6463.07	
2. Validation sample (N=273)	4444.84	2948	1.51	.93	.043	.041-.046	.080	5028.84	
3. All sample (N=818)	6944.33	2948	2.36	.96	.041	.039-.042	.064	7528.33	
4. All sample (N=818): Keep paths with $\beta \geqq .05$ only	**6952.59**	**2977**	**2.34**	**.96**	**.040**	**.039-.042**	**.064**	**7478.59**	
5. Boy sample (N=429)	5565.88	2906	1.92	.94	.046	.044-.048	.075	6073.88	
Free PIQ→Anxiety	5555.27	2905	1.91	.94	.046	.044-.048	.075	6065.27	
Free VIQ→Anxiety	5532.87	2904	1.91	.94	.046	.044-.048	.074	6044.87	
Free VIQ→ Others' stereotype	5498.30	2903	1.89	.94	.046	.044-.048	.074	6012.30	
Free Spatial ability→ Anxiety	**5480.12**	**2902**	**1.89**	**.94**	**.046**	**.044-.047**	**.074**	**5996.12**	
6. Girl sample (N=389)	**4725.16**	**2906**	**1.63**	**.95**	**.040**	**.038-.042**	**.069**	**5233.16**	

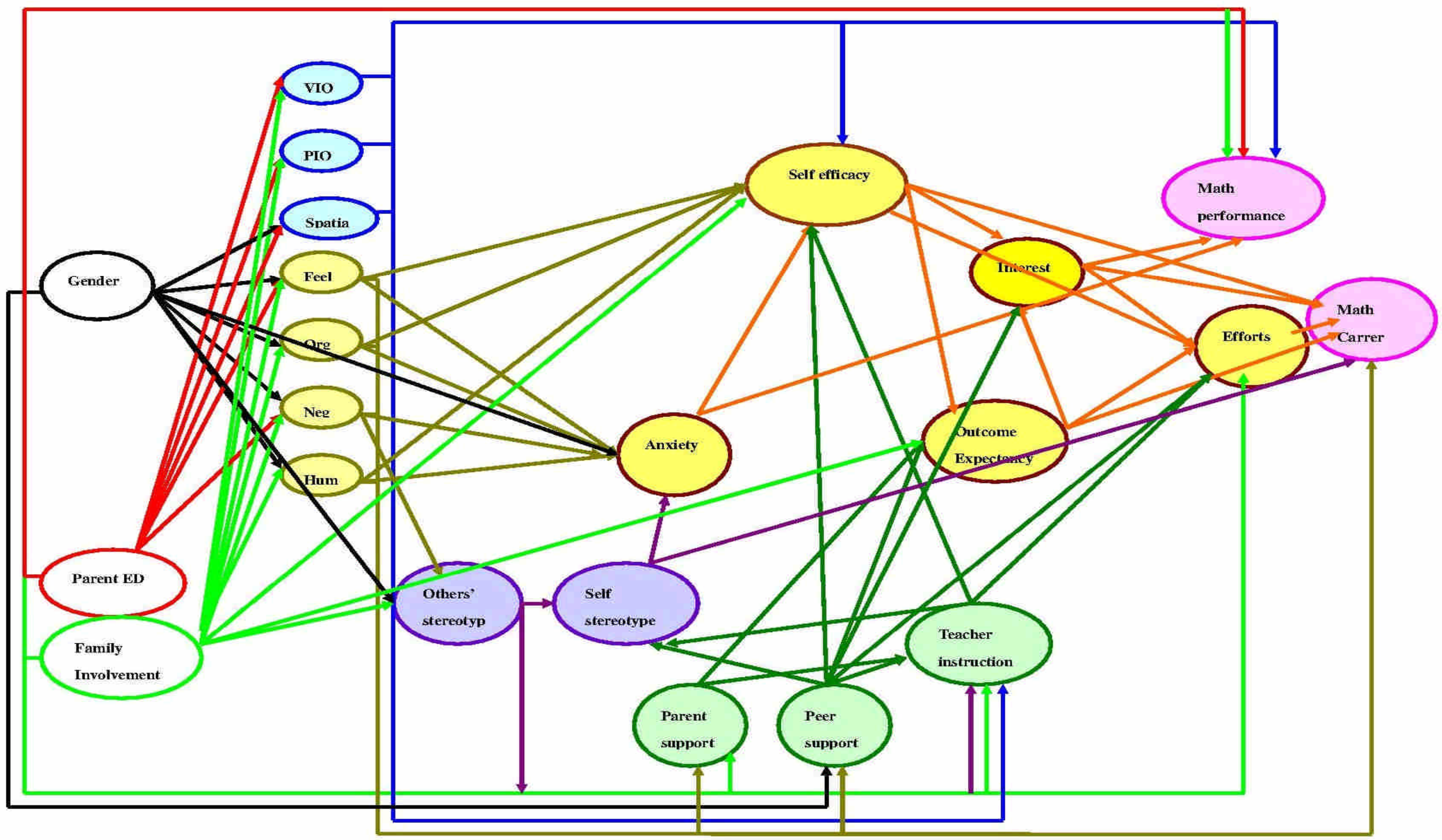


Figure 2a. Final model ($\beta \geqq .05$) for entire Elementary 4th grade sample (N=818):

Table 2b. Hypotheses testing for the Junior high school sample (Grade 7):

model	χ^2	df	χ^2/df	CFI	RMSEA	RMSEA 90%CI	SRMR	AIC	aBIC
1. Calibration sample (N=740)- starting	8002.83	2984	2.68	.94	.048	.046-.049	.070	8514.83	
1b. Calibration sample (N=740)-modified	6596.97	2958	2.23	.96	.041	.039-.042	.060	7160.97	
2. Validation sample (N=362)	4829.36	2958	1.63	.95	.042	.040-.044	.070	5393.36	
3. All sample (N=1102)	7919.58	2958	2.68	.96	.039	.038-.040	.057	8483.58	
4. All sample (N=1102): Keep paths with $\beta \geqq .05$ only	**7940.95**	**2991**	**2.65**	**.96**	**.039**	**.038-.040**	**.058**	**8438.95**	
5. Boy sample (N=568)	5923.46	2920	2.03	.95	.043	.041-.044	.068	6403.46	
Free Peer→ Interest	5888.62	2919	2.02	.95	.042	.041-.044	.068	6370.62	
Free Stereotype→Self Efficacy	5858.57	2918	2.01	.96	.042	.041-.044	.067	6342.57	
Free teacher quality→ parents math support	5846.68	2917	2.00	.96	.042	.041-.044	.067	6332.68	
Free human-style→ outcome expectancy	5822.95	2916	2.00	.96	.042	.040-.043	.067	6310.95	
Free Self-efficacy→ outcome expectancy	5811.45	2915	1.99	.96	.042	.040-.043	.067	6301.45	
Free feeling-style→ peer math support	**5793.52**	**2914**	**1.99**	.96	**.042**	.040-.043	.067	**6285.52**	
6. Girl sample (N=534)	5215.80	2920	1.79	.96	.038	.037-.040	.064	5695.80	
Free Negative style → Outcome Expectancy	5194.17	2919	1.78	.96	.038	.037-.040	.064	5676.17	
Free Spatial ability → Anxiety	**5180.99**	**2918**	**1.78**	**.96**	**.038**	**.036-.040**	**.063**	**5664.99**	

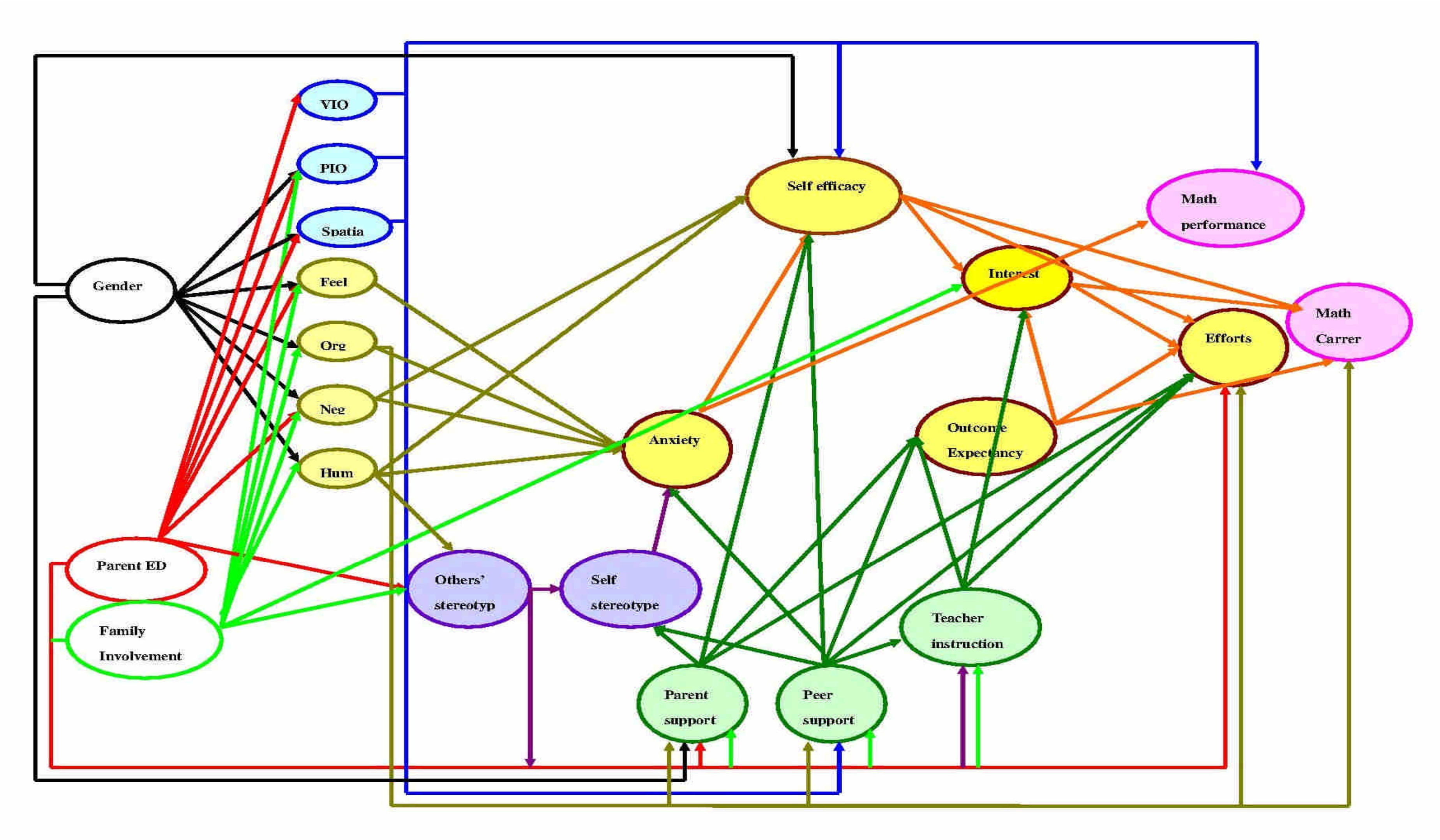


Figure 2b. Final model ($\beta \geqq .05$) for entire Junior high school 7th grade sample (N=1,102):

Table 2c. Hypotheses testing for the Senior high school sample (Grade 10):

model	χ^2	df	χ^2/df	CFI	RMSEA	RMSEA 90%CI	SRMR	AIC	aBIC
1. Calibration sample (N=825)- starting	9188.09	2984	3.08	.93	.050	.049-.051	.072	9700.09	
1b. Calibration sample (N=825)-modified	6753.81	2944	2.29	.95	.040	.038-.041	.055	7345.81	
2. Validation sample (N=412)	4923.00	2944	1.67	.94	.040	.038-.042	.066	5515.00	
3. All sample (N=1237)	8410.28	2944	2.86	.96	.039	.038-.040	.052	9002.28	
4. All sample (N=1237): Keep paths with $\beta \geqq .05$ only	**8434.76**	**2973**	**2.84**	**.96**	**.039**	**.038-.040**	**.053**	**8968.76**	
5. Boy sample (N=717)	**6378.74**	**2906**	**2.20**	**.95**	**.041**	**.039-.042**	**.057**	**6886.74**	
6. Girl sample (N=520)	4991.89	2906	1.72	.95	.037	**.035-.039**	.063	5499.89	
Free teacher quality→ self stereotype	**4969.52**	**2905**	**1.71**	**.95**	**.037**	**.035-.039**	**.061**	**5479.52**	
Free peer support→ Perceived others' stereotype	**4963.15**	**2904**	**1.71**	**.95**	**.037**	**.035-.039**	**.059**	**5475.15**	

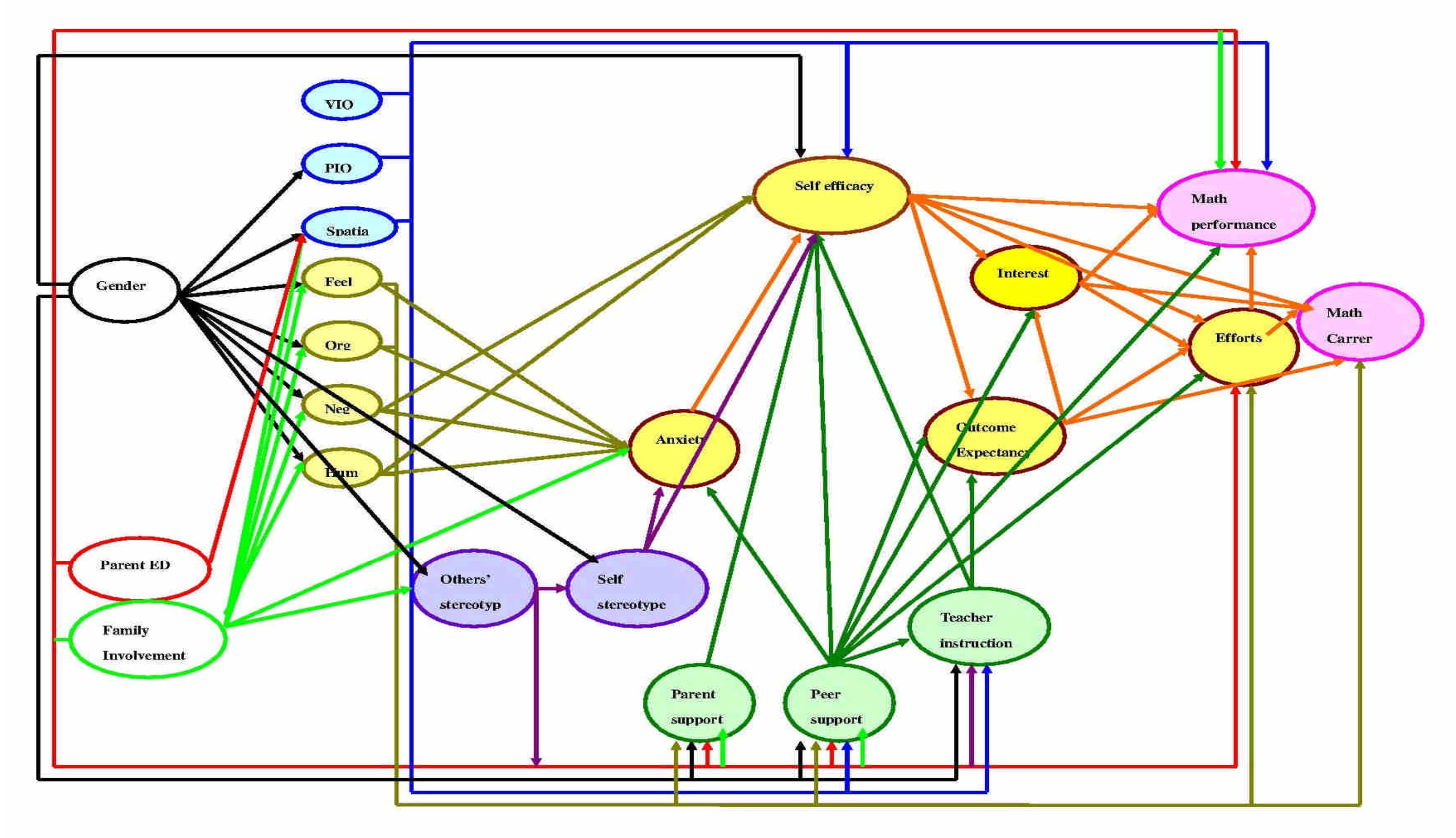


Figure 2c. Final model ($\beta \geqq .05$) for entire Senior high school 10th grade sample (N=1,237):

3. Effect of 'gender' on students' mathematics learning

Based on the previously identified models (models shown in figure 2a, 2b, and 2c), the effect of gender on other variables were further investigated. The direct and total effect of gender on other variables were reported in Table 3.

(3.1) Effect of 'gender' on 'mathematic performance' and 'math career intention'

'Math performance' and 'math career intention' were the main interests of this research. They have been treated as the core dependent variables, and the main goal of this study was to find a comprehensive model which best explains their variations. Thus, investigate how gender affects these two important dependent variables are fundamental in this research.

As data shown in table 3, for children in all three age levels, gender does not have a direct effect on either math performance or math career intention. This is an important finding, because it means that all the observed gender differences on these two dependent variables were all mediated by other intervening factors. This finding also supported the need to dig into a complex model, like what we have done here, in order to find the ultimate answer for the gender-math paradox.

The total effect of gender on math performance is .06, .13, and .15 for grade 4, 7, and 10, accordingly. The total effect of gender on math career intention is .02, .14, and .22 for grade 4, 7, and 10, accordingly. The effect of gender seems getting larger as age grows older (noticed also that our 10^{th} grade sample is from senior high school, which is also a group with higher abilities). For junior and senior high school adolescents, gender generally shows a moderate effect on their math performance and career intention. The influence of gender on career intention seems a slightly larger than its influence on math performance. On the average. Boys show higher scores on math exam from their

Table 3. Direct and total effect[1] (in parenthesis) of gender on all studied variables[1,2]:

DV	Elementary (N=818) 4th grader	Junior (N=1,102) 7th grader	Senior[3] (N=1,237) 11th grader
1. Verbal IQ	---(---)	--- (---)	--- (---)
2. Performane IQ	---(---)	.12 (.12)	.08 (.08)
3. Spatial ability	.08(.08)	.07 (.07)	.08 (.08)
4. Feeling style	**-.42 (-.42)**	**-.39 (-.39)**	**-.29 (-.29)**
5.Organized style	**-.31 (-.31)**	**-.19 (-.19)**	**-.15 (-.15)**
6. Negative style	**-.14 (-.14)**	**-.21 (-.21)**	**-.16 (-.16)**
7. Human style	**-.21 (-.21)**	**-.13 (-.13)**	**-.10 (-.10)**
8. Others' stereotype	**.27 (.23)**	--- (.04)	**.35 (.35)**
9. Self stereotype	**--- (.15)**	--- (.01)	-.14 (.06)
10. Anxiety	**-.20 (-.10)**	**--- (-.16)**	**--- (-.08)**
11. Parents' support	--- (-.07)	.05 (.01)	.06 (.02)
12. Peers' support	**.28 (-.13)**	**--- (-.12)**	**.07 (-.04)**
13. Qualityof Instruction	**--- (-.13)**	--- (-.05)	.05 (-.05)
14. Self efficacy	**--- (.13)**	**.19 (.24)**	**.25(.30)**
15. Outcome expectancy	--- (-.01)	--- (-.05)	--- (.01)
16. Interest	--- (.06)	**--- (.18)**	**--- (.20)**
17. Efforts	**--- (-.08)**	**--- (-.10)**	**--- (-.12)**
18. **Math performance**	--- (.06)	**--- (.13)**	**--- (.15)**
19. **Career intention**	--- (.02)	**--- (.14)**	**--- (.22)**

Note 1. Gender code: girl=0, boy=1
Note 2. unstandardized and standardized values are the same
Note 3. senior group is a higher ability group with a smaller variation

elemantary years, and they started to show higher math-related career intention in their adolescent time. The observed gender differences are getting larger as age grows up.

(3.2) Effect of 'gender' on other variables

For the elementary 4th graders, major findings according to table 3 were: (1) large effect: gender shows the largest total effect on personal styles. Boys are less feeling-oriented (-.42), and have a tendency to be less organized (-.31), or in the other words, more flexible; (2) moderate effect: gender shows moderate total effects(.10-.24) on several other variables, such as negative style, human-oriented style, perceived others stereotype, self stereotype, math anxiety, perceived environmental support from peers, perceived math teachers' quality, and math self efficacy; (3) small effect: gender has small effect(.05-.09) on perceived parents' support, spatial ability, math interest, and math efforts; (4) no effect: while taking all variables into consideration in the same frame, gender shows no significant effect on VIQ, PIQ, and math outcome expectancy.

Altogether, for children who are at 9 years old, the general gender similarity and gender differences are as this: Compared to girls, 9-year-old boys show similar VIQ and PIQ average, while slightly higher spatial ability. However, they have a very different personal style: they are less feeling- and human-oriented, less organized, and less negative-oriented. With these basic background differences, boys at this age report to perceive stronger gender-math stereotype from others; Boys themselves also believe that male generally perform better in the math-related field. Ironically, even boys feel somewhat less math-support from environment (teachers, peers, and parents), and make less effort in studying mathematics, they still show less math anxiety, higher math-self efficacy, and higher math interest. Notice also that boys at this age do show slightly higher math performance, while they have not stronger math-related career intention at this stage.

For the junior high school 7th graders, major findings according to table 3 were: (1) large effect: gender shows the largest total effect on personal styles. Boys are less feeling-oriented (-.39); (2) moderate effect: gender shows moderate total effects(.10-.24) on several other variables, such as PIQ, organized style, negative style, human-oriented style, math anxiety, perceived environmental support from peers, math self-efficacy, math interest, and math effort; (3) small effect: gender has small effect(.05-.09) on spatial ability, perceived teacher quality, and outcome expectancy; (4) no effect: while taking all variables into consideration at the same time, gender shows no significant effect on VIQ, perceived others stereotype, self stereotype, and perceived parents' support.

Altogether, for adolescents who are at 12 years old, the general gender similarity and gender differences are as this: Compared to girls, 12-year-old boys show similar VIQ average, but have higher spatial and PIQ abilities. It means that boys may better handle nonverbal information in learning. Meanwhile, Boys continue to reveal a very different personal style: they remained to be less feeling- and human-oriented, less organized, and less negative-oriented. With these basic differences, it is interest to find that gender differences on perceived/ and self math-gender stereotype diminished somewhat at this stage (We later checked data and found it was due to the magnitude of girls' perception of stereotype increased at this stage). While gender show similar feeling about support from parents, boys continue to feel less math support from peers and teachers, and to make less effort in mathematics. However, they continue to show less math anxiety, higher math-self efficacy, and higher math interest (these tendency is even stronger compared to the 4th graders). Notice also that adolescent boys at this age do show higher math performance, and their intention to pursue math-related career is also getting stronger significantly.

For the senior high school 10th graders (also a higher ability group overall), major findings according to table 3 were: (1) large effect: gender shows the largest total effect on perceived others' stereotype. Boys feel others showing more math-gender stereotype (.35). Meanwhile, boys' math-self-efficacy is getting stronger(.30), while they maintained to be less feeling-oriented (-.29); (2) moderate effect: gender shows moderate total effects(.10-.24) on several other variables, such as organized style, negative style, human-oriented style, math interest, and math effort; (3) small effect: gender has small effect(.05-.09) on PIQ, spatial ability, self math-gender stereotype, math anxiety, and perceived teachers' quality; (4) no effect: while taking all variables into consideration at the same time, gender shows no significant effect on VIQ, perceived support from parents and peers, and math outcome expectancy.

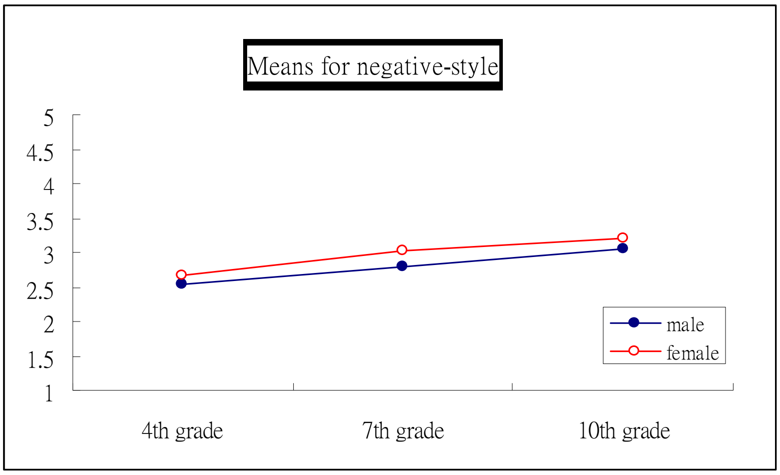
Altogether, for adolescents who are at 15 years old, the general gender similarity and gender differences are as this: Compared to girls, 15-year-old boys show similar VIQ average, but have higher spatial and PIQ ability. Meanwhile, they continued to reveal a very different personal style: they remained to be less feeling- and human-oriented, less organized, and less negative-oriented. With these basic differences, boys themselves believe that male is better at math, and they perceived strongly from others about this stereotype. Boys continue to feel less math support from teachers and make less effort in math. Even though, they continue to show less math anxiety, salient stronger math-self efficacy, and higher math interest. Notice also that adolescent boys at this age show even better math performance and stronger math-related career intention.

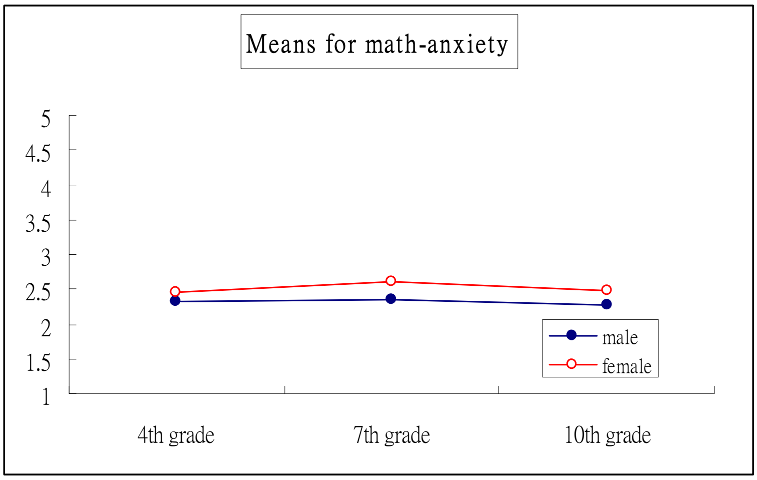
From a developmental perspective, some interesting trends of gender-similarity and gender differences are noticed: (1) For all three age levels, both genders are similar at the average VIQ and math outcome expectancy. It means that both genders can handle verbal information with similar proficiency, and both groups value the mathematics the same; (2) Boys constantly perform better on spatial ability with the same level of magnitude from age 9 to 15, and they start to show higher non-verbal ability (PIQ) in adolescent years; (3) Gender difference on personal style is salient and stable across age bands. However, the magnitude of gender gap on many of these characteristics (such as feeling-oriented, organized-style, and human-oriented style) are getting smaller as age grows older. The one exception was the style of ‘negative-oriented’, girls show constant higher level of negative style, and the gender gap reaches the largest at age 12 ; (4) Boys at all three age levels feel less support from peers and teachers; while both genders feel similar level of parents support since age 12; (5) Boys across all ages show lower math anxiety, higher math interest, and higher math-efficacy. As age grows up, gender difference on math interest and self-efficacy is getting larger; (6) Boys across all ages make less effort on math study. As age grows up, gender gap on paid efforts is also getting larger;

Besides, an especially interesting observed trend was this, gender difference on math anxiety and negative style reach the largest level at 7th grade, this is also the time where gender gap on gender-math stereotype reach the lowest level! Isn’t it surprising?

We made a detailed checking of the data, and the means for each gender were presented in Figure 3. Generally, results show the followings: As age grows older, negative-style for both genders are getting stronger. 7th grade girls show a salient jump on negative style, yielding the biggest gender gap at this level. Data also revealed that, boys do not show big variation on math anxiety across different age levels, however, 7th grade girl show a salient jump on math anxiety, yielding the biggest gender gap at this level. Besides, boys tend to perceive more stereotype from others in general. As age grows older, both gender perceive more and more math-gender stereotype from others. However, from grade 4 to grade 7, boys’ perception do not change much, but girls’ perception jump up a lot. That is the reason for finding the smallest gender gap on perceived stereotype at this age. In Erikson’s theory, self-identity is the main crisis for adolescents at this age (age 12). Physiologically, it is the time gender hormone functions the most. Girls at this age show a jump on level of negative-style, they also explicitly perceive more gender-math stereotype from others at this time point. At the same time, boys' advantage on spatial and nonverbal abilities also become more salient. Summing together, grade 7 is an important stage, it is the time girls show a jump on math anxiety, and on their perception of others gender-math stereotype.

While data in Table 3 revealed a direct effect of gender on math self-efficacy as .00, .19, and .25 for grade 4, 7, and 10(the total effect is .13, .24, and .30 correspondingly), this rising trend made us wondering that the variable "gender" itself may be qualitatively different somewhat along the developmental trend. Math self-efficacy may have been embedded into the deep inside of "self", at some time point between age 9 to 12. If, this truly is the critical time for this gender-math self efficacy bond to form, it also should be treated as the critical time for effective educational intervention, if possible.





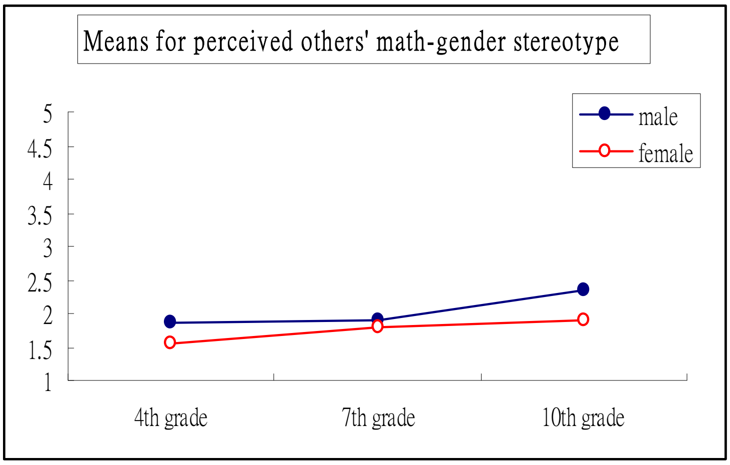


Figure 3. Means for each gender on negative-style, math anxiety, and perceived other stereotype

4. Effect of all studied variables on 'math perfomance' and 'math career intention':

Based on models generated for each gender across the three age levels, Table 4 presented the direct and total effects of all studied variables on the two major dependent variables, math performance and math career intention, for each gender across the three age levels. Several important findings are listed below.

(4.1) Important and stable predictive factors for math performance and math career intention

As expected, the factors which show the largest total effect on 'Math career intention' and 'Math performance' are distinctly different!

For Math career intention, the factors which show moderate-large effect for both genders across all three age levels are: math interest, math self-efficacy, math anxiety, peer support, math outcome expectancy, and organized-oriented style. While for both genders, the top predictive factors for math performance at all ages are ability factors (PIQ, VIQ, and spatial ability). Table 5 presented top predictive variables for each gender in each age level. It is clearly shown that cognitive abilities are the core factor for academic math performance; while the other conative factors (such as interest, perceived support from peers, perceived self-efficacy, personal style, anxiety, and outcome expectancy) are the fundamental dominant sources for students' decision on pursuing a math-related career.

Most importantly, personal style factors (such as feeling-oriented style, organized style, negative style, and human-oriented style) show small to large effect on career intention, but their effects on math performance are trivial. On the contrary, cognitive ability factors (such as VIQ, PIQ, and spatial ability) show strong effects on math performance (total effects across all samples are between .21 to .74), but their effects on math career intention is significantly lower (total effects for 4th grades are between .02 to .15; for 7th graders are between .00 to .08; for 10th graders are between .03 to .26). Results clearly revealed that performance and career intention are quite different constructs, they need to be investigated separately.

(4.2) Stable gender differences found across all three age levels:

Based on table 4, results revealed some stable gender differences. First of all, stereotype was found to show a stronger impact for girls, especially on math career intention. For girls, perceived gender-math stereotype from others has moderate effect on math career intention, while self gender-math stereotype has moderate to large effects on this same variable. For boys, most of these effects are smaller. Consequently, with the same magnitude of perceived gender-math stereotype, girls' intention for math career drop down more rapidly than boys do.

Second, math self-efficacy shows large effect on math career intention for both genders, However, the impact for girls is relatively larger. It means that with one standard deviation increase in math self-efficacy, we predict the math-career intention for girls could be improved more.

(4.3) Observed variations across three age levels:

Based on table 4, as age grows up, some variations across age band deserve attention:

First of all, the importance of family involvement/ support on math career intention decreases at senior high school. The total effects are .44, .33, and .01 for girls; while they were .44, .31, and .09 for boys (noticing that this could due to the higher ability of our grade 10 sample).

Second, math interest is found to show large effect on career intention for both genders across all age levels. However, the relative importance of this variable decreases slightly as age grows up. The total effects are 1.45, .83, and .73 for girls; while they were 1.28, 1.08, and .90 for boys.

Third, the influence of math anxiety on math career intention is getting larger too. The total effect for girls groups are -.29, -.36, and -.41; for boys, they are -.14, -.35, and -.43. While anxiety blocks the intention, the impact reaches the largest at grade 10.

Fourth, compared to the other two younger samples, the importance of nonverbal abilities(such as PIQ and spatial ability) on career intention become much larger for 10^{th} graders. Simarly, the importance of self-efficacy, math interest, and math efforts on math performance all become much larger for 10^{th} graders. Again, it could be that we are using a higher ability senior-high school sample in this research. It deserves further examination.

Finally, according to Table 4, it is clearly shown that math self efficacy tends to affect math performance in different patterns for students at different ages. For 4^{th} graders, the higher the self-efficacy, the lower the math performance; for 7^{th} graders, these two constructs do not relate; for 10^{th} graders, the higher the self-efficacy, the higher the math performance. Similarly, math interest also tends to affect math performance in different patterns for students with different ages. For 4^{th} graders, the higher the math interest, the lower the math performance; for 7^{th} graders, these two constructs do not relate; for 10^{th} graders, the higher the math interest, the higher the math performance. What could the underlying mechanism for this trend? Is it possible the perception of self ability for 4^{th} graders was influenced by some factors other than their previous math performance? Is it possible that for 4^{th} graders, ability and interest constructs are somewhat more distinctive to each other, and less embroiled in their mind? It deserves more works for exploration.

able 4. Direct and total effect[1] (in parenthesis) of all variables on Math Performance and Math career intention:

IV \ group	DV= Math Performance						DV=Math Career Intention					
	Grade 4		Grade 7		Grade 10[2]		Grade 4		Grade 7		Grade 10[2]	
	boy	girl	boy	girl	boy	girl	boy	girl	boy	girl	boy	girl
1. PED	-.11 (.08)	-.13 (.05)	--- (.32)	--- (.20)	-.09 (.04)	-.10 (-.04)	--- (.02)	---(.04)	--- (.02)	--- (.04)	--- (.01)	--- (.00)
2. Family support	-.08 (.04)	-.03 (.11)	---(.02)	---(.09)	.15 (.02)	.08 (.02)	--- (.44)	--- (.44)	--- (.31)	--- (.33)	--- (.09)	--- (.01)
1. Verbal IQ	.56 (.55)	.60 (.60)	.60 (.60)	.70 (.70)	.35 (.34)	.21 (.21)	--- (.15)	--- (.02)	--- (.00)	--- (.00)	--- (.10)	--- (.03)
2. Performane IQ	.71 (.69)	.70 (.68)	.69 (.69)	.74 (.74)	.36 (.37)	.25 (.29)	--- (.15)	--- (.09)	--- (.06)	--- (.06)	--- (.16)	--- (.17)
3. Spatial ability	.40 (.40)	.53 (.51)	.44 (.44)	.47 (.48)	.51 (.52)	.37 (.44)	--- (.06)	--- (.09)	--- (.08)	--- (.07)	--- (.26)	--- (.20)
4. Feeling style	--- (-.04)	--- (.04)	--- (-.01)	--- (-.01)	--- (-.04)	--- (.02)	--- (.42)	--- (.33)	-.09 (-.04)	-.08 (-.14)	-.06 (.10)	-.09 (.05)
5.Organized style	--- (-.02)	--- (-.05)	--- (.01)	--- (.00)	--- (.06)	--- (.11)	-.09 (.27)	-.09 (.36)	.17 (.38)	.03 (.22)	-.04 (.09)	-.02 (.18)
6. Negative style	--- (-.03)	--- (-.09)	--- (-.02)	--- (-.03)	--- (-.02)	--- (-.07)	.08 (-.02)	.04 (-.13)	--- (-.07)	--- (-.10)	--- (-.19)	--- (-.19)
7. Human style	--- (.00)	--- (-.05)	--- (.00)	--- (.01)	--- (-.05)	--- (-.04)	-.02 (-.02)	-.15 (-.15)	--- (.11)	--- (.01)	--- (.05)	--- (.04)
8. Others' stereotype	--- (-.01)	--- (-.06)	--- (-.01)	--- (-.01)	--- (.02)	--- (-.05)	--- (-.07)	--- (-.15)	--- (-.03)	--- (-.11)	--- (.04)	--- (-.12)
9. Self stereotype	---(-.01)	---(-.09)	--- (-.01)	--- (-.03)	--- (.03)	--- (-.11)	-.10 (-.13)	-.12 (-.26)	--- (-.02)	--- (-.12)	--- (.06)	--- (-.24)
10. Anxiety	-.09 (-.05)	-.25 (-.18)	-.03(-.03)	-.05 (-.05)	--- (-.19)	---(-.18)	--- (-.14)	--- (-.29)	---(-.35)	---(-.36)	---(-.43)	---(-.41)
11. Parents' support	--- (.00)	--- (.00)	--- (.00)	--- (.00)	--- (.01)	--- (.01)	---(.01)	---(.03)	.02(.02)	.05 (.05)	--- (.06)	--- (.07)
12. Peers' support	--- (-.05)	--- (-.05)	--- (.04)	--- (.00)	-.37 (-.07)	-.21 (-.03)	--- (.69)	--- (.60)	--- (.49)	--- (.51)	--- (.64)	--- (.42)
13. Qualityof Instruction	--- (.02)	--- (.01)	--- (.00)	--- (.00)	--- (-.02)	--- (.01)	--- (-.01)	--- (-.08)	---(.08)	---(.16)	--- (-.02)	--- (.03)
14. Self efficacy	-.09 (-.12)	-.08 (-.14)	--- (.00)	--- (-.00)	.17 (.28)	.21 (.31)	-.44 (.45)	-.55 (.56)	-.32 (.48)	-.02 (.62)	.02(.63)	.16(.70)
15. Outcome expectancy	--- (.00)	--- (-.01)	--- (.00)	--- (.00)	--- (.08)	--- (.04)	.07 (.17)	.15 (.34)	.06 (.13)	.16 (.23)	.12 (.31)	.18 (.26)
16. Interest	-.05 (-.05)	-.07 (-.07)	--- (.00)	--- (.00)	.20 (.25)	.11 (.11)	1.28 (1.28)	1.42 (1.45)	1.08 (1.08)	.83 (.83)	.89 (.90)	.73 (.73)
17. Efforts	.02 (.02)	-.01 (-.01)	--- (.00)	--- (.00)	.17 (.17)	.21 (.21)	.09 (.09)	.13 (.13)	--- (.00)	--- (.00)	.02 (.02)	.09 (.09)

ote1. unstandardized and standardized values are the same Note 2. Grade 10 is a group with higher ability

Table 5. Top predictive variables (total effect) for math performance and math career intention:

	Math Performance						Math Career intention					
	Grade 4		Grade 7		Grade 10[1]		Grade 4		Grade 7		Grade 10[1]	
Rank	**boy**	**girl**	**boy**	**girl**	**boy**	**girl**	**boy**	**girl**	**boy**	**girl**	**boy**	**girl**
1	PIQ (.69)	PIQ (.68)	PIQ (.69)	PIQ (.74)	Spatial ability (.52)	Spatial ability (.44)	Interest (1.28)	Interest (1.45)	Interest (1.08)	Interest (.83)	Interest (.90)	Interest (.73)
2	VIQ (.55)	VIQ (.60)	VIQ (.60)	VIQ (.70)	PIQ (.37)	Self-efficacy (.31)	Peer support (.69)	Peer support (.60)	Peer support (.49)	Self-efficacy (.62)	Peer support (.64)	Self-efficacy (.70)
3	Spatial ability (.40)	Spatial Ability (.51)	Spatial ability (.44)	Spatial ability (.48)	VIQ (.34)	PIQ (.29)	Self-efficacy (.45)	Self-efficacy (.56)	Self-efficacy (.48)	Peer support (.51)	Self-efficacy (.63)	Peer support(.42) Anxiety (-.38)
4	Self-efficacy (-.12)	Anxiety (-.18)	Parents' education (.32)	Parents' education (.20)	Self-efficacy (.28)	VIQ (.21)	Family support (.44)	Family support (.44)	Organized-Style (.38)	Anxiety (-.36)	Anxiety (-.43)	Anxiety (-.41)
5	Parents' education (.08)	Self-efficacy (-.14)	---	Family support (.09)	Interest (.25)	Efforts (.21)	Feeling-style (.42)	Organized-Style (.36)	Anxiety (-.35)	Family support (.33)	Outcome expectancy (.31)	Outcome expectancy (.26)
Others variables which show large effects	---	---	---	---		---	Organized-Style (.27)	Outcome expectancy (.34) Feeling-style (.33) Anxiety (-.29) Self-stereotype (-.26)	Family support (.31)	Outcome expectancy (.23) Organized-Style (.22)	Spatial ability (.26)	

ote 1. Grade 10 is a group with higher ability

5. Peeking into the inside of the complex models: linking paths with large direct effects:

Beyond the results presented in the previous sessions, a remained interesting question might be "what are the main linking paths in the model?" or "Is there any gender difference on the major linking paths?" To better answer these questions, we presented figures of the simplified models, one for each age level, with only large direct effects were shown ($\beta \geqq .25$). Figure 4a, 4b, and 4c presented these information for grade 4, 7, and 10, correspondly. According to information from these 3 figures, main findings are as followings:

(5.1) strong impact paths which are stable for both genders across all 3 age levels:

First of all, as described earlier, ability factors (especially nonverbal abilities such as PIQ and spatial ability) have large direct effect on math performance for both genders across all age levels. Second, The more support form the family (parents care more about students overall learning), the more organized-style the child is. Third, a very salient and stable chain of linking paths for all groups is 『negative-styled➔math anxiety➔math self-efficacy➔math interest➔math career intention』. Individual different attributes like personal style (negative-oriented) does show strong impact on math learning, for students with various age levels, also with different genders. Fundamental individual differences should not be ignored in any plan/form of education intervention. Fourth, Perceived stereotype from others affects the student's self math-gender related stereotype.

(5.2) Strong paths which are for all boys only

First of all, VIQ has a large and direct effect on math performance for boys at all ages. Second, family support has a direct, large, and stable effect on boys' personal styles. The more general learning involvement and support from the family, the more feeling- and human-oriented the boy is. Third, the more feeling-oriented the boy is, the more support he feels from the peers. Fourth, the more peer support boys feel, the better teacher quality they perceived, the higher self-efficacy they have, and the more efforts they make for math study.

(5.3) Strong paths which are for all girls only

First of all, the more organized-style the girls are, the more support they feel from peers. Second, the higher the gender-math stereotype the girls themselves have, the higher the level of their math anxiety.

(5.4) Other findings based on overall model:

First, when checking the overall total effects in the whole model, the organized style showed salient association with many other math-related factors (such as math efforts, math interest, outcome expectancy, self efficacy, and perceived peer support). This might be one personal style which education can expect to make a change gradually.

Second, if considering the total effects altogether, a very salient chain is 『perceived others' stereotype→ self stereotype → math anxiety → math self-efficacy → math interest → math career intention』. Generally, children with more negative and less human-oriented styles tend to feel others showing much stronger math stereotype.

Third, when considered the total effects altogether, family background is again proved to be a fundamental core for childrens' lives. Its' impact is everywhere. Students from a more supportive home environment is found to show more organized and more humanistic style, they are less likely to view others with math stereotype, they feel less math-anxiety, feel more support from environment (parents, peers, and teachers),have higher self-efficacy and willing to pay more efforts.

Forth, compared to support from parents and teachers, perceived peer support is found to be a stronger factor for students' overall math learning. This variable is related to many other factors such as perceived teacher quality, self efficacy, outcome expectancy, math interest, math efforts, and math career intention. Results suggested that how to well utilize the power of peer-cooperation in students' math learning is definitely a direction for math instructors to proceed.

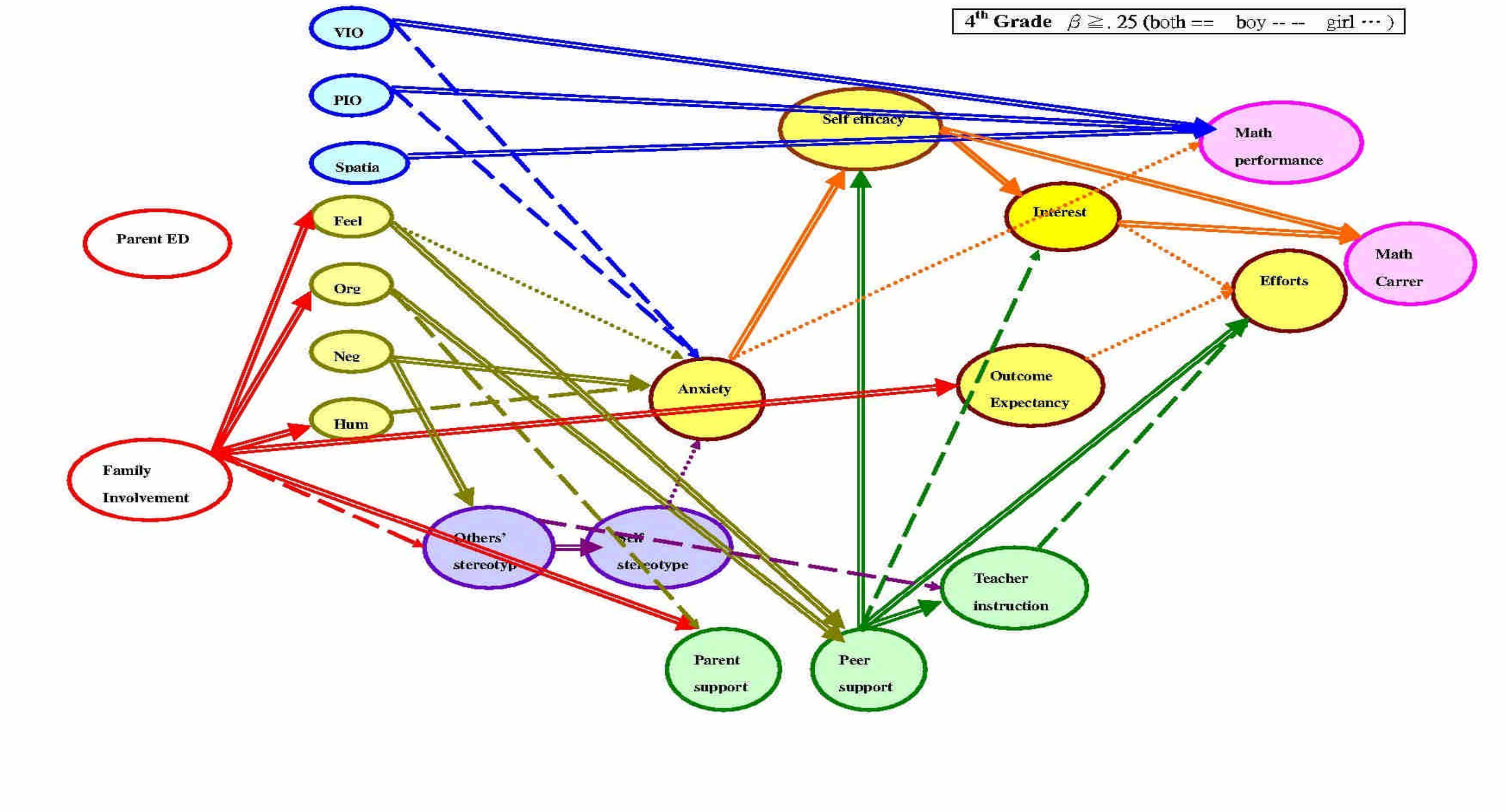


Figure 4a. Paths with large direct effect ($\beta \geqq .25$) for 4th grade boys and girls:

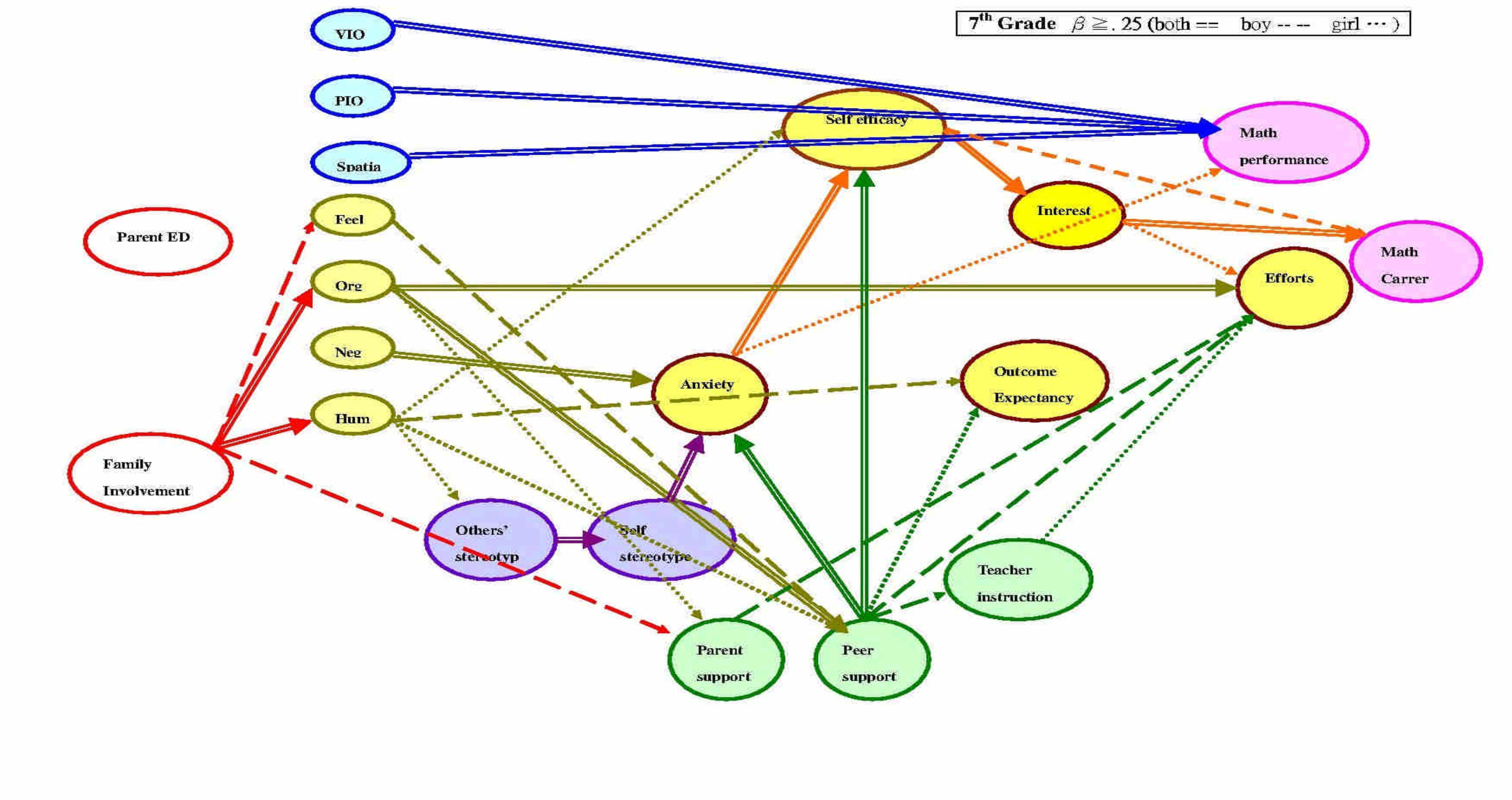


Figure 4b. Paths with large direct effect ($\beta \geqq .25$) for 7th grade boys and girls:

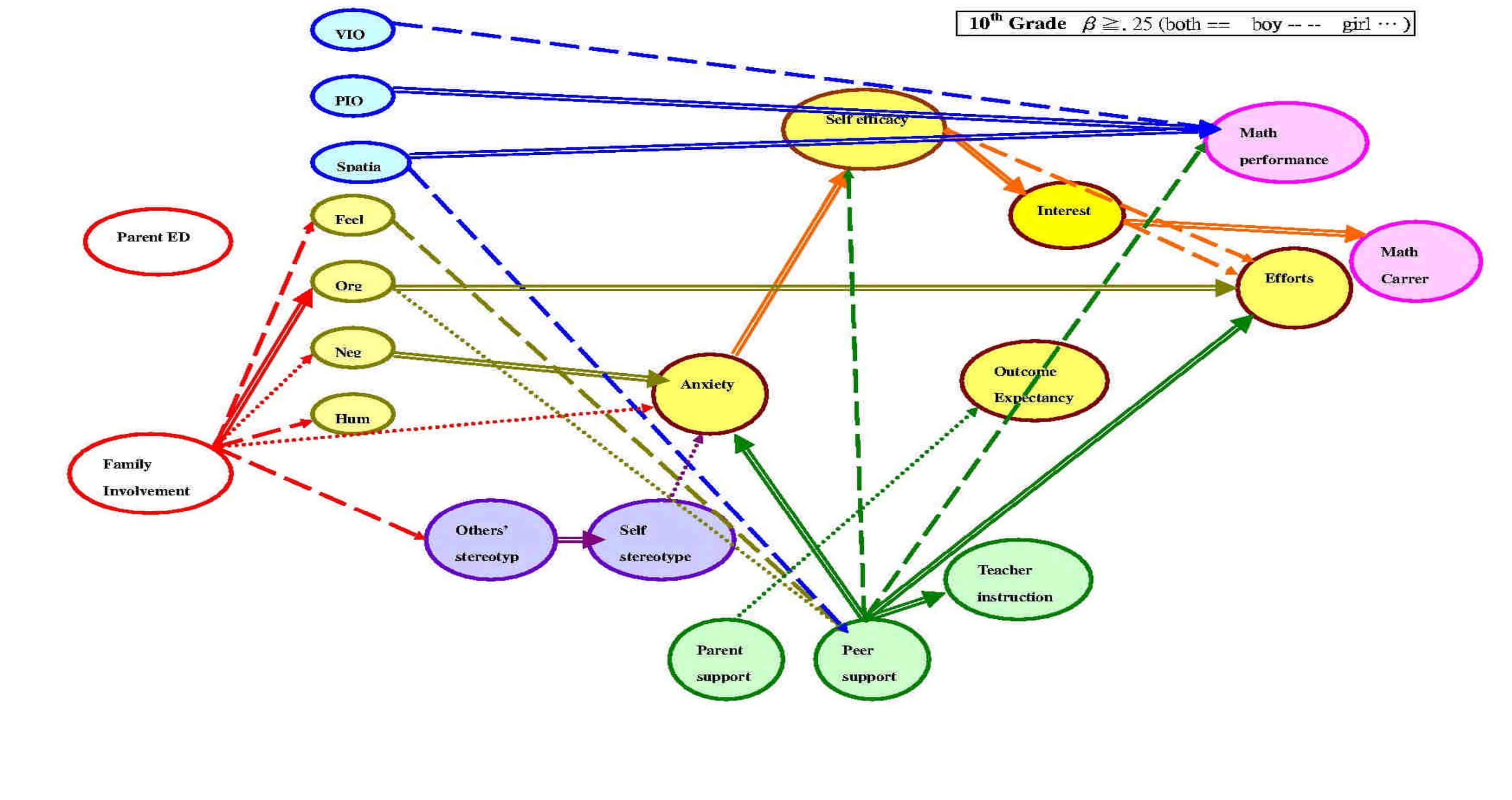


Figure 4c. Paths with large direct effect ($\beta \geqq .25$) for 10th grade boys and girls:

Discussion

Several interest gender-similarity and gender-differences were revealed in our research. The uniqueness of this study were (1) a comprehensive and important list of variables were selected and jointly investigated, (2) data based on a comprehensive age levels (grade 4, 7, and 10) provided valuable developmental trend for us to observe, and (3) each sample is followed for three years (results to be reported in later papers). This study thus was considered able to make an important contribution for the field of gender-math research .

In this research, our models explain data well. Moreover, we were able to explain large portion of the variances of important dependent variables. For example, our models explain roughly over 90% of the math performance variance for 4^{th} and 7^{th} graders; while it was about 43-61% for the 10^{th} graders (noticing that our 10^{th} graders are a group of higher ability students). Besides, current models explain over 90% of the math career intention variance for 4^{th} graders (it was due to the high correlation between math interest and math career intention), 81-84% of the total variance for the 7^{th} graders, and 88-92% of the total variance for the 10^{th} graders.

The major findings of this research are:

(1) The fundamental learning model is generally the same for both genders from age 9 to 15. This can be proved by the results of goodness-to–fit index for the starting model reported in the result section. Thus the questions to be answered mainly would be " whether these constructs had the same effect across genders? " or "how different the magnitude of influence of these constructs are for different gender? ".

(2) All effects of gender on math performance and math career intention are indirect, but not direct! Furthermore, the top predictive factors for both constructs are quite different!

This is an important finding because it shows that all the impact of gender on the two important dependent variables were mediated by other intervening variables in childrens' lives. The total effect of gender on math performance is .06, .13, and .15 for grade 4, 7, and 10, accordingly. The total effect of gender on math career intention is .02, .14, and .22, accordingly. Effect of gender seems getting larger as age grows up.

Factors which show the largest total effect on 'Math career intention' and 'Math performance' are distinctively different! For math career intention, the factors which show moderate-large effect for both genders across all age levels are: math interest, math self-efficacy, math anxiety, outcome expectancy, organized style, and peer support. While the top predictive factors for math performance for all students are ability factors such as PIQ, VIQ, and Spatial ability.

Personal style factors (such as feeling-oriented style, organized style, negative style, and human-oriented style) show small to large effect on career intention, but their effect on math performance is a lot smaller. On the contrary, ability factors (such as VIQ, PIQ, and spatial ability) show significantly large effects on math performance (total effects across all samples are between .21 to .74), but their effects on math career intention is significantly lower (total effects for 4^{th} grades are between .02 to .15; for 7^{th} graders are between .00 to .06; for 10^{th} graders are between .03 to .26). Results clearly revealed that performance and career intention are two different constructs, the estimated relationship between these two latent constructs were .15, .13, and .42 for grade 4, 7, and 10.accordingly. Thus math performance and math career intention need to be improved and nurtured through different paths.

(3) The way boys and girls differ varies as age grows.

At grade 4 (age 9), boys show similar mean VIQ and PIQ to girls, while have slightly higher spatial ability. However, gender differences on personal style are large and significant: boys are less feeling- and human-oriented, less organized, and less negative-oriented. With these basic differences, boys at this age perceive more gender-math stereotype from others also themselves. They feel less math support from environment and make less effort in math, However, boys show less math anxiety, higher math-self efficacy, and higher math interest. Notice also that boys at this age do show slightly higher math performance, while they have no stronger math-related career intention.

At grade 7 (age 12), boys have higher spatial and PIQ ability. Meanwhile, patterns of large gender differences on personal style remain unchanged (although the gap is slightly smaller). Interestingly, although boys continue to feel less support from peers and teachers, also make less effort in math, they continue to show less math anxiety, higher math-self efficacy, and higher math interest. The gender gap on perceived stereotype actually diminished a bit because of the stereotype feeling gilrs have is jump up. Notice also that adolescent boys at this age do show higher math performance, and their intention to pursue math-related career is also getting stronger significantly.

At grade 10 (age 15), boys keep to show higher spatial ability and PIQ. Patterns of large gender differences on personal style remain unchanged (while the gap is getting even smaller). Boys continue to feel less math support from teachers and make less effort in math. However, boys at this age feel stronger math-gender stereotype from others and themselves. They continue to show less math anxiety, higher math-self efficacy, higher math interest, higher math performance, and math-related career intention.

Gender difference on math anxiety and negative style reach the largest level at 7th grade, but it also is the time where gender gap on perceived stereotype reaches the lowest level. The reasons came from 7th grade girls show a salient jump on negative style, math anxiety, and perceive gender-math stereotype from others. It is assumed that from somewhere between age 9 to 12, it is critical point for girls to form the math-anxiety association. More intervention should be focus on the age band, for expecting getting more effective results.

Another important issue to notice was that, a direct effect of gender on math self-efficacy was found to be .00, .19, and .25 for grade 4, 7, and 10 correspondingly. This rising trend made us wondering that the variable "gender" itself may be qualitatively different along the developmental trend. Math self-efficacy may have been embedded into the deep core of "self", at some time point between age 9 to 12. If, this is the critical time for this bond to form, it also should be treated as the critical time for needed educational intervention.

(4) Family support is important!

Family background is proved to be a fundamental source of students' lives. Its impact is everywhere. Students from a more general learning supportive home environment is found to show more organized and more humanistic style, they are less likely to view others with math stereotype, they feel less math-anxiety, they feel more support from environment (parents, peers, and teachers), have higher self-efficacy and willing to pay more efforts. Parents-education should be encouraged.

(5) Personal styles are important!

When the total effects are considered, organized-style is significantly associated with many more math-related factors (such as career intention, math efforts, math interest, outcome expectancy, and self efficacy); besides, for both genders, the ones with more negative and less human-oriented styles tend to feel others showing much stronger math stereotype. .

Although these personal-style constructs are usually considered as background variables, and are known not be changed easily. However, paying more attention in nurturing the appropriate personal styles, or educating students explicitly about how to deal with negative emotions, should be a continuous effort for all educators.

(6) Make students feel being supported by peers in their math class, is important!

Compared to math support from parents and teachers, perceived peer support is found to be the more important environmental factor on students overall math learning. Educators should try to design more peer-interacted program in math class, create opportunities for students feel a supportive atmosphere from peers. This, in tern, will have higher chance making his/her math learning better improved.

(7) As age grows older, the importance of family support on career intention decreases, the importance of abilities (such as PIQ and spatial) and self-efficacy on career intention increases. Math interest is found to show the largest effect on career intention for both genders across all age levels. However, the relative importance of this variable decreases slightly. Besides maintaining students' interest, how to empower students and help them feel really able is a continuous important action.

(8) For adolescents in all three age levels, math stereotype, and math self-efficacy show comparatively stronger impacts for girls. These constructs should be paid more attention in intervention project which aims at improving girls' performance and intention in math /science related field

(9) The observed strong and stable direct paths for both genders were: (a) nonverbal ability factors (PIQ and spatial ability) have large direct effect on math performance. (b)The more parents care about students overall learning, the more organized-style the child is. (c) The higher the math anxiety, the lower the math self-efficacy, the lower the math interest, thus the lower the math career intention. Individual different attributes like personal style does show strong impact on math learning. (d) the more the perceived stereotype from others, the more self math-gender stereotype.

The Strong paths in boys' group only are 『VIQ→ math performance』, 『family support→ Perceived stereotype 』, 『feeling oriented→perceived peer support』, 『perceived peer support→ perceived teacher quality』 , 『perceived peer support→ self efficacy』,and 『perceived peer support→ math efforts』. Obviously peer support is especially an important factor for boys' math learning overall.

The Strong paths in boys' group only are 『organized style→ perceived peer support』, 『perceived stereotype from others→ self stereotype』, girls feeling tend to be influenced by others more easily.

Overall, our research provides rich information about gender similarities and differences on mathematics learning. Nonetheless, there is still inevitable limitations. Due to practical concern, the ability level for our three samples varied somewhat. As reported in the method section. The mean VIQ of the 7th grade sample is more close to the population mean, thus can be considered as a more representative sample. The other two samples (grade 4 and grade 10) are higher ability groups with roughly 0.5 SD higher on verbal intelligence. Readers should notice this characteristic and apply current findings carefully.

As Halpern, Benbow, Geary, Gur, Hyde, and Gernsbacher (2007, p.41) wonderfully concluded, “There is no single factor by itself that has been shown to determine sex differences in science and math. Early experience, biological constraints, educational policy, and cultural context each have effects, and these effects add and interact in complex and sometimes unpredictable ways”. Our findings help revealing how complex the connections between the math learning-related constructs are. Webb, Lubinski, & Benbow (2002) suggested equal gender representation across all educational-vocational domains may conflict with what might be happening naturally. Thus, “equal male-female representation across disciplines may not be as simple to accomplish as many policy discussions imply (Webb et al. 2002, p.785)”. We believe, the progress of science lies in better knowing what can be done and what might not be easily changed. It would help if researchers and policy makers could view the gender-math paradox with a scientific and objective mind.

References (略)

Appendix 1. Percentages of each cognitive and content domains for self-developed math exams

Cognitive domain	Elementary G3-G6	Junior-high G7-G9	Senior-high G10-G12
Knowing	40%	35%	30%
Applying	40%	40%	40%
Reasoning	20%	25%	30%

	Elementary				Junior-high			Senior-high		
Content domain	**G3**	**G4**	**G5**	**G6**	**G7**	**G8**	**G9**	**G10**	**G11**	**G12**
Number	36%	34%	34%	40%	46%	13%	0%	30%	0%	0%
Measurement	29%	19%	20%	12%	5%	0%	0%			
Geometry	21%	28%	23%	24%	0%	59%	39%	40%	60%	25%
Algebra	7%	13%	14%	20%	49%	28%	25%	30%	10%	48%
Data and Chance	7%	6%	9%	4%	0%	0%	36%	0%	30%	27%
overall	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

計畫成果自評：

本研究在團隊成員的用心努力下，得到豐富的資料與重要的發現！研究內容與原計畫相符！我們不僅前半段研究中，完整文獻閱讀並整理國內大型測驗結果，找出男女學生學習相關主要差異所在。 本計劃三年研究之前半部份，已於 2010 年發表一篇 SSCI 期刊

Chen, H., Chen, M, Chang, T., Lee, Y & Chen, H. P. (2010). Gender Reality on Multi-domains of School-age Children in Taiwan: A Developmental Approach. *Personality and Individual Differences,* 48(4), 475-480. (SSCI)

之後，我們根據前述發現,慎選重要變因進入下一階段研究，成功建立不同年齡,不同性別學生共六個數學學習預測模式。成果已報告如前，有眾多有趣發現。

在國內性別數學相關研究中,我們的研究是第一篇能包含如此多元向度變項(推理能力,空間關係能力，數學成就表現，及與學習意動相關題目,共有 80 個項目組成 22 個潛在因素,整體平均信度在 0.80 上下)，又能包含如此完整年齡階段(小學,中學,高中)的大型研究(正式研究樣本為 3,157 位 9-15 歲學生)。這樣的 scope 即便在國際研究中也不多見！因此我們對研究品質與成果有信心！研究成果不僅具有學術價值,對數學教育實務界也能有幫助!

出席國際學術會議心得報告

計畫編號	NSC 96-2522-S-003-017-MY3
計畫名稱	影響不同性別學生數學學習相關因素之發展趨勢,與可能模式之建立與驗證（第 3 年）
出國人員姓名 服務機關及職稱	陳心怡 國立台灣師範大學特殊教育學系教授
會議時間地點	San Diego, CA, USA　Aug 12-15, 2010
會議名稱	the 118th Annual Convention of the American Psychological Association (APA)
發表論文題目	Chen, H., Hung, L., Huang, Y., Chen, H., Cheng, S., & Wong, S. (2010). Interventions for students with special needs in Taiwan: A quantitative synthesis of single-subject researches.

一、參加會議經過 與心得

今年的 Annual Convention of the American Psychological Association (APA)在美國聖地牙哥舉辦，會期由 8/12-15 共四天。這個會議是美國最大型的心理學會議，會議內容包含心理學各個領域的研究，參與人員也來自世界各地！讓我們有機會與世界各地研究者互動討論。此外,除了學術性研討外，展場也有多達上百所出版心理學相關書籍,工具,測驗,或儀器之出版社參加展覽,因此,可讓我們在短短幾天內提升心理學領域最新相關知識與訊息！會議型式多元，包括 Keynote address、Symposium、Paper present, Poster Sessions、與 Roundtable discussion 等等。

連續四天會議有上千件的學術報告， 在心理學眾多領域中,我主要對特殊兒童教育,以及認知與測驗評量相關議題有興趣,此外由於目前在進行的研究是分析學生數學學習相關因素之發展趨勢,與可能模式之建立與驗證。故選擇參與的多場 Keynote address 與 Symposium 多與認知與學習等此類領域有關。 此外， 我也在 paper 與 poster session 中與國外研究者進行交流與互動討論。 為期數天的短暫充電過程感到收獲甚多。

舉例而言， 我參與的其中一場 symposium， 題目是 ”The next decade of motivation research- an interactive discussion”。五位與談學者針對當前動機研究的各不同理論(如

achievement goal theory, self-determination theory 等)提出看法,會議中對於不同理論內所涵蓋的 constructs 重疊的現況進行討論,並分析未來如何加以適當統整或許是學界所樂見之發展方向。

此外,與學習相關的其他座談中,如自我效能(self-efficacy)重要性,執行功能(executive functioning)與工作記憶(working memory)等議題, 動機情緒間聯結,及提升學生主動學習的教學策略, 以及情緒控制策略,社會技巧介入方案之設計等議題也都被大家所重視討論。而與測驗編製技術相關的研討也不少,有的場此討論如何改變計分方式以強化構念的測量,也有議題是與測驗編製過程中的常模建立問題有關,例如由 Dr. Roid 與 Dr. Gorsuch 兩位擔任主講,所討論 continuous norming 的新發展與應用便很值得學習。此一概念已被許多當代主要大型測驗所採用,這些交流加強了我對相關議題瞭解與掌握之能力。

觀摩他人新近研究成果,以及與各國同領域學者進行面對面意見交換討論,應是參與國際性學術會議最大的收穫! 經由觀摩別人的發表可以瞭解不同領域的最新走向;而由深入討論中則可充份吸收與交換不同學者的多元觀點, 瞭解不同地區的獨特發展與考量,這些都有助增進研究者本身學術思考的完整度。

近幾年來,隨著一次次國際研討會的參與,我由每次的與會都再次真實感受到一個學術研究者視野深度與廣度長期養成的重要! 當培養出更上一層的視野,面對研究問題所產生的思緒將隨之改變! 我逐漸瞭解到學術研究前輩為何常常退而不修. 因為,每當進入新一層次的高度,又是一個新面向挑戰與學習的開始! 長期實力的累積過程,只有一步一腳印, 沒有捷徑! 在學習過成中,研究的有趣性,挑戰性,與成就感更吸引我往前邁進。

衷心感謝國家對我們學術研究者的培育, 我們有機會登高望遠, 隨著個人研究層次與實力的提升,我們將有更多的力量貢獻與回饋!

二、報告摘要

Interventions for students with special needs in Taiwan: A quantitative synthesis of single-subject researches

Chen, H., Hung, L., Huang, Y., Chen, H., Cheng, S., & Wong, S.

I Objectives:

The population of students with special needs is emerging, which in turn raised questions about the best instructional intervention for remediating learning problems for these children. Effective intervention is essential in the field of special education (Gersten et al., 2000; Kavale,1990). Several meta analyses have been reported for synthesize the empirical evidence for exceptional childrens (Forness, Kavale, Blum, & Lloyd,1997; Jones, 2005; Lloyd, Forness,& Kavale, 1998; Swanson, 1999, 2006; Swanson & Deshler, 2003; Swanson, Hoskyn, & Lee, 1999; Swanson & Lee, 2000). From a cross-cultural perspective, the purpose of this study was to enrich current understanding by synthesizing the effectiveness of intervention outcomes for students with special needs in Taiwan.

II Methods:

A total of 270 single-subject studies, which produced 2128 effect sizes, were acquired by a through search of major Taiwan journals published in 1995 to 2007. All studies met the following criteria: (1) The objective was to investigate the effectiveness of special education interventions for exceptional students; (2) A single-subject research design, with either reversal or multiple-baseline design was employed. Studies with A-B design were excluded; (3) Baseline and treatment phases were presented clearly for individual participants, which provided enough information for calculating effect size.

Coded information covered seven major domains: (1) Basic characteristics for the study (ex. year of publication, publish status); (2) Characteristics for the participant (ex. disability type, participant age, educational setting.); (3) Characteristics for the research design (ex. reversal or

multiple-baseline design, research quality); (4) Characteristics for the intervention (ex. duration of intervention, background knowledge of the interveners, instructional strategies.); (5) Major treatment (Independent) variable ; (6) Major outcome (dependent) variable ; and (7) quantitative information .

For over 3 years of devoting, four doctoral students majoring in special education served as coders. By frequent and intensive discussions, disagreements were resolved in regular group meetings, and the overall inter-coder reliability was approaching .91. For each baseline-treatment pair, both the PND (percentage of non-overlapping data) and PEM (percentage of data points exceeding the median of baseline phase) (Ma, 2006) procedures were employed to compute effect size.

III Results and discussion:

The grand mean PND was .79 (SD=.31), and the grand mean PEM was .87 (SD=.23). The lag 1 autocorrelation was found to be significant, thus indicating the violation of assumption of independency. Nonparametric statistics thus were applied for following analyses. According to Scruggs et al (1986) and Ma (2006), the average PND and PEM scores, which were between .70 and .90, revealed a moderate effect.

Further works on moderator identification were based on results of Kruskal-Wallis test and Mann-Whitney U test. The findings depicted that the magnitude of effect size varies significantly across various levels of publication year, disability type, students' age, student's IQ, research quality, duration of training, and background knowledge of the interveners. Regarding to effect of major treatments, compared to computer-assisted instruction and writing intervention, remediation on word recognition, math, and reading comprehension showed somewhat smaller effects.

Besides, for a total of 34 investigated instructional strategies, relatively better strategies for various type of exceptional students were identified. Briefly speaking, the strategies with larger effect sizes for students with learning disabilities were (1) adopt self-management approach such as self-evaluation, self-monitoring and self-reinforcement; (2) arrange appropriate setting or environment for learning; (3) design class on a one-to one basis; and (4) define intervention goal clearly. The better instructional strategies for emotional/behavioral disorders were (1) prompt students to use strategies or procedures; (2) involve parents in the learning process; (3) direct students to pay attention on what is being taught; and (4) define intervention goal clearly. The better strategies for teaching mental retarded children were (1) adjust intervention goals flexibly during teaching process, (2) utilize token system; (3) arrange appropriate setting or environment for learning ; (4) adopt self-management approach such as self-evaluation, self-monitoring and self-reinforcement; and (5) monitor the progressions of students' learning. The better strategies for teaching autistic students were (1) offer either oral of written feedback; (2) use token system; (3) use help from volunteers; (4) reinforce desired behaviors; and (5) adjust intervention goals flexibly during teaching process.

These abovementioned findings provided evidences for better understanding the effectiveness of intervention outcomes for students with special needs. This information can be compared with findings from other nations for providing cross-cultural comparisons.

國科會補助計畫衍生研發成果推廣資料表

日期:2011/10/28

國科會補助計畫	計畫名稱: 影響不同性別學生數學學習相關因素之發展趨勢,與可能模式之建立與驗證
	計畫主持人: 陳心怡
	計畫編號: 96-2522-S-003-017-MY3　　學門領域: 數學教育—科學教育理論—數學

無研發成果推廣資料

96 年度專題研究計畫研究成果彙整表

計畫主持人：陳心怡			計畫編號：96-2522-S-003-017-MY3				
計畫名稱：影響不同性別學生數學學習相關因素之發展趨勢,與可能模式之建立與驗證							
成果項目			量化			單位	備註（質化說明：如數個計畫共同成果、成果列為該期刊之封面故事...等）
			實際已達成數（被接受或已發表）	預期總達成數(含實際已達成數)	本計畫實際貢獻百分比		
國內	論文著作	期刊論文	0	0	100%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	0	0	100%		
		專書	0	0	100%		
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力（本國籍）	碩士生	0	0	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		
國外	論文著作	期刊論文	1	0	100%	篇	Chen, H., Chen, M, Chang, T., Lee, Y & Chen, H. P. (2010). Gender Reality on Multi-domains of School-age Children in Taiwan: A Developmental Approach. Personality and Individual Differences, 48(4), 475-480. (SSCI)
		研究報告/技術報告	0	0	100%		
		研討會論文	0	0	100%		
		專書	0	0	100%	章/本	
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	

		權利金	0	0	100%	千元	
	參與計畫人力（外國籍）	碩士生	0	0	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		

其他成果 (無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等，請以文字敘述填列。)	目前研究成果已發表一篇 SSCI 期刊 Chen, H., Chen, M, Chang, T., Lee, Y & Chen, H. P. (2010). Gender Reality on Multi-domains of School-age Children in Taiwan: A Developmental Approach. Personality and Individual Differences, 48(4), 475-480. (SSCI) 研究小組也正進行後續論文撰寫,準備再發表於 SSCI 期刊中

	成果項目	量化	名稱或內容性質簡述
科教處計畫加填項目	測驗工具(含質性與量性)	0	
	課程/模組	0	
	電腦及網路系統或工具	0	
	教材	0	
	舉辦之活動/競賽	0	
	研討會/工作坊	0	
	電子報、網站	0	
	計畫成果推廣之參與（閱聽）人數	0	

國科會補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

■達成目標

□未達成目標（請說明，以 100 字為限）

　　□實驗失敗

　　□因故實驗中斷

　　□其他原因

說明：

2. 研究成果在學術期刊發表或申請專利等情形：

論文：■已發表 □未發表之文稿 □撰寫中 □無

專利：□已獲得 □申請中 ■無

技轉：□已技轉 □洽談中 ■無

其他：（以 100 字為限）

本研究之初步成果已有一篇發表在國際 SSCI 期刊中

Chen, H., Chen, M, Chang, T., Lee, Y & Chen, H. P. (2010). Gender Reality on Multi-domains of School-age Children in Taiwan: A Developmental Approach. Personality and Individual Differences, 48(4), 475-480. (SSCI)

目前仍持續整理後續資料,將會再發表論文

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）（以 500 字為限）

本研究的價值在於由全面性發展角度,來檢視不同年齡,不同性別學生的數學學習!

本研究在團隊成員的用心努力下,得到豐富的資料與重要的發現! 我們不僅前半段研究中, 完整文獻閱讀並整理國內大型測驗結果, 找出男女學生學習相關主要差異所在。 本計劃三年研究之前半部份, 已於 2010 年發表一篇 SSCI 期刊

Chen, H., Chen, M, Chang, T., Lee, Y & Chen, H. P. (2010). Gender Reality on Multi-domains of School-age Children in Taiwan: A Developmental Approach. Personality and Individual Differences, 48(4), 475-480. (SSCI)

在第二階段研究中,我們慎選 22 個與男學學生數學學習有關的重要變因, 成功建立國小, 國中,高中三個不同年段,不同性別學生共六個數學學習預測模式。能有效解釋台灣學生的數學成就與數學生涯選擇意願之變異!成果至為豐碩!

在國內性別數學相關研究中,本研究是第一篇能包含如此多元向度變項(推理能力,空間關係能力, 數學成就表現, 及與學習意動相關題目,共有 80 個項目組成 22 個潛在因素,整體

平均信度在 0.80 上下)，又能包含如此完整年齡階段(小學,中學,高中)的大型研究(正式研究樣本為 3,157 位 9-15 歲學生)。這樣的 scope 即便在國際研究中也不多見！因此我們對研究品質與成果有信心！研究成果不僅具有學術價值,對數學教育實務界也能提出具體建議！