Why Girls Do Better in Mathematics in Hawai‘i: A Causal Model of Gender Differences on Selected and Constructed Response Items

American Educational Research Association

San Francisco
April 07 - 11, 2006

Patrica P. Reiss

University of Hawai‘i at Manoa
preiss@hawaii.edu

Shuqiang Zhang

University of Hawai‘i at Manoa
szhang@hawaii.edu

Why Girls Do Better in Mathematics in Hawaiʻi: A Causal Model of Gender Differences on Selected and Constructed Response Items

Research evidence has consistently shown that female students in Hawaiʻi outperform males in mathematics. Brandon, Newton, and Hammond (1987), who examined data from the 1982 and 1983 mathematics Stanford Achievement Test (SAT) administered to Hawaiʻi public school students in grades four, six, eight, and ten, found that overall, females consistently outperformed males across these grade levels. Brandon and Jordan (1994) examined the 1991 SAT mathematics results for tenth graders in Hawaiʻi and confirmed that girls performed better than boys.

Brandon and Jordan (1994) also analyzed the results for grade eight on the mathematics section of the 1990 National Assessment of Educational Progress (NAEP) and found that "of the 40 participating jurisdictions, Hawaiʻi was the only one in which girls' total-test mean scores were significantly higher than boys'" (p. 18). The results of the 2000 administration of the NAEP for grades four and eight also show that in Hawaiʻi, "unlike national results," females scored higher than males in mathematics (Hawaiʻi State Department of Education, 2001), with the same pattern reported for the 2003 administration (Hawaiʻi State Department of Education, 2003).

What these studies fail to elucidate, however, is why girls in Hawaiʻi consistently outperform boys in mathematics. Nor do the published studies make clear to what extent the gender-related difference in mathematics is attributable to gender-related differences in reading and writing. An extensive literature search failed to reveal any Hawaiʻi-based study that investigated the unique females' advantage over males in mathematics in conjunction with linguistic factors. To the best of the authors' knowledge, this paper represents the first research attempt on the topic and presents potential pedagogical implications or policy adjustment necessary to optimize mathematics education for boys and girls.

Literature Review

The female advantage in mathematics in Hawai‘i lies in sharp contrast to the findings in the continental U.S. Ten large-scale mainland-based U.S. studies involving at least 1,000 students each were identified and reviewed. For overall mathematics performance, Cole (1997), Nowell and Hedges (1998), Wilson and Zhang (1998), and the Office of Educational Accountability (2002) all found that males outperformed females. For constructed-response (CR) items, however, the findings are inconclusive. A majority of studies found that females perform better than males (DeMars 1998, 2000, Garner & Engelhard, 1999; Myerberg, 1996; Zhang & Manon, 2000). For selected-response (MC) items, the consistent finding is that males perform better than females (DeMars, 1997, 1998, 2000; Garner & Engelhard, 1999, Myerberg, 1996; Wilson & Zhang, 1998; Zhang & Manon, 2000).

Table 1 about here

To gain a better understanding of how gender differences may vary from one item format to another, a meta-analysis was undertaken (see Table 1) encompassing 15 independent MC and 14 independent CR gender difference effect sizes collected from four studies that were published between 1998 and 2000, and reported boys’ and girls’ scores separately on MC and CR items. Despite the fact that both MC and CR items are widely adopted in large-scale mathematics assessment nowadays, the rather limited number of published studies on how the two formats may differentially affect males and females points to a clear need for further research in the area. No format-related effect sizes are available from any of the Hawai‘i based published studies.

Table 2, which summarizes effect sizes on MC items, shows a consistent, though small, gender difference in favor of males. Hedges’ g, an effect size indicator for the standardized differences between means, was obtained by dividing the mean difference between the female and male means by the pooled standard deviation (Hedges & Olkin, 1985; Rosenthal, 1991). The test of heterogeneity of the effect sizes is non-significant ($\chi^2_{(14)} = 22.64$, $p > 0.05$). The average size of the effect is $g = -0.06$, indicating that the score of the average male on MC mathematics tests surpasses that of 52.39% of females.

Tables 2 & 3 about here

For the CR format, the results are inconsistent but favor females in 60% of the individual effect sizes examined (see Table 3). The test of heterogeneity for the CR effect sizes is significant ($\chi^2_{(13)} = 211.01$, $p < 0.05$). Even though the mean effect size ($g = +0.01$) is in favor of females, this should be interpreted with caution because the effect sizes are heterogeneous. It is obvious from Tables 2 and 3 that gender-differences in mathematics vary across formats.

That females are advantaged on CR items might be explainable by females' superior performance on tests of language ability. There is a general consensus that females perform better than males on tests of verbal ability (Cole, 1997; Halpern, 2000, 2004; Hyde & Linn, 1988; Maccoby & Jacklin, 1974; Nowell & Hedges, 1998). Coley (2001), for example, reported that the results of the 1992, 1994, and 1998 NAEP for grades 4, 8, and 12 showed that females outperformed males in reading and writing for all racial/ethnic groups. Writing or verbal ability has been suggested as a possible reason for the better performance of females as compared to males on constructed-response mathematics items (Willingham & Cole, 1997).

Studies focusing on the impact of verbal skills on mathematics performance of students whose native language is not English have been conducted by a number of researchers (Abedi, 2000; Abedi, Lord, & Plummer, 1995; De Avila, 1988; Kiplinger, Haug, & Abedi, 2000). The results of these studies strongly suggest that language ability influences mathematics scores. One such study was conducted in Hawaiʻi by Gronna, Chin-Chance, and Abedi (2000), who concluded that "English language proficiency affects students assessment scores on standardized norm referenced tests in the content areas of reading and Mathematics *[sic]*" (p. 9). However, their study provides no information on how language proficiency may affect performances on MC and CR items respectively. Therefore, a study to examine the mathematics section of the Hawaiʻi State Assessment (HSA) program, with a particular focus on establishing a causal model that accounts for performances on MC and CR items by considering the effects of gender and language ability, would seem both relevant and necessary.

This study, based upon the 2002 HSA data for third and fifth graders, attempts to confirm a causal model that would explain how gender, verbal skills, and mathematics performance on MC and CR items are causally connected. Such a study might eventually assist the state of Hawaiʻi in developing gender-appropriate intervention to adjust current mathematics education and meet the requirements of the No Child Left Behind Act (NCLB).

A Causal Model

The model (see Figure 1) includes one exogenous variable: gender; and four endogenous variables: writing score, reading score, mathematics MC score, and mathematics CR score. (In the HSA data set, gender is coded 1 = female, 0 = male.) Of the four endogenous variables, three also serve as mediating variables, writing score, reading score, and mathematics MC score, which pass various indirect gender effects on to the ultimate endogenous variable, CR score.

Figure 1 about here

The continental U.S. studies reviewed show males to have an advantage over females on tests of mathematics, whereas the Hawai‘i data show females to have an advantage over males. Figure 1 depicts a causal model that takes into account the hypothesized better performance of males on tests of mathematics ability as compared to females, as well as the hypothesized better performance of females on tests of verbal ability as compared to males.

The paths from gender to MC and CR in Figure 1 can be predicted to be either positive or negative. Given the recurring female advantage in mathematics in Hawai‘i over the last two decades, a positive direct effect from gender to CR or MC is conceivable. However, given the consistent male advantage in the continental U.S. when overall mathematics and the MC format are considered, a negative direct effect from gender to MC is also plausible. Based on the review of the literature and theoretical understanding, it is difficult to predict the direction of the path from gender to CR. Only 8 out of the 14 effect sizes examined above pointed to a female advantage on CR items.

The paths from gender to reading and writing are predicted to be positive because females have been shown nearly always to perform better on tests of verbal ability. Reading is shown in Figure 1 to have a positive effect on MC and CR for females. Similarly, writing should have a positive effect on CR and reading. Writing influences the reading variable because the HSA reading test incorporates CR items, which necessitate a written response.

In addition to the direct effects of gender on MC and CR, this model explains mathematics performance by considering the various paths that involve mediating variables. For example, the indirect effect of gender on MC is made up of the path that leads from gender via writing via reading to MC, and the path that leads from gender via reading to MC. The overall gender effect is a synthesis of both direct and indirect effects, illustrating how males and females may follow quite different direct and indirect paths to arrive at their respective performance levels.

Methods

Instruments. The HSA, designed specifically according to the revised Hawai‘i Content and Performance Standards (HCPS II) (Hawai‘i State Department of Education, 2005), assessed mathematics, reading and writing in 2002. The reading and mathematics sections for all grades tested included MC and CR items. For mathematics, two types of CR items, short- and extended response, were combined to make up a CR score. The writing section required students to produce an essay. (For sample items see Hawai‘i Department of Education, 2004).

Sample. The data set used in this study included 6,352 girls and 6,354 boys in the third grade, and 6,331 girls and 6,717 boys in the fifth grade. Students who received alternate tests or special accommodations were excluded from the analysis. Also excluded were students for whom one or more scores on the reading, writing, or mathematics tests were missing, or whose gender had not been identified.

Variables. For each grade, the standards-based reading, writing, MC mathematics, and CR mathematics scores were calculated. Descriptive statistics for the variables are given in Table 4. Cronbach's alpha, given in Table 5, ranges from 0.83 to 0.96, which shows that the items are highly inter-related and that the tests have satisfactory internal consistency.

Tables 4 and 5 about here

Analyses. For each grade, a separate path analysis was performed to investigate the proposed model. Because gender is coded dichotomously, only unstandardized path coefficients are reported. Analyses were based on the variance-covariance matrices derived from the raw scores for all variables (See Tables 6 and 7). The path analysis was conducted using the SAS System's CALIS procedure.

Tables 6 and 7 about here

Results

The results of the path analyses are reported in Figures 2 and 3 for the third and fifth grades respectively. One interesting finding, consistent in both grades, is that the direct paths from gender to both mathematics-MC and mathematics-CR are in favor of males despite the well-documented findings in Hawai‘i that girls outperform boys in overall mathematics performance (Brandon et al., 1987; Brandon & Jordan, 1994;

Hawai‘i State Department of Education, 2001; 2003). This suggests that it is too simplistic to examine mathematics scores without taking into consideration the effect of language factors.

Figures 2 and 3 about here

The goodness of fit of the model was determined through the chi-square test and other commonly employed goodness of fit indices (see Table 8). The chi-square test with one degree of freedom is statistically significant, $\chi_{(1)}^2$ = 221.82 at grade three and $\chi_{(1)}^2$ = 172.01 at grade five, $p < 0.01$. Because of the very large sample size, it was expected that the outcome of the chi-square test would be significant. The other goodness of fit indices afford support of the model’s fit at both grade levels. The GFI, AGFI, CFI, NNI, and NFI amount to at least 0.9 providing support for the model’s fit at both grade levels.

Table 8 about here

The nine path coefficients in the model for grade three and five are shown in Figures 2 and 3 respectively. A positive path coefficient indicates that the path favors females, while a negative path coefficient specifies an advantage for males. All path coefficients presented are statistically significant ($p < 0.01$) except for the path coefficient from gender to reading, which is non-significant ($p > 0.05$) at grade three.

The direct effects show that the path leading from gender to CR is negative, and the path leading from gender to MC is also negative. Examining the path coefficient from gender to CR, one observes that for females, the CR score decreases by -0.54 points for grade three and –0.30 points for grade five when all other independent variables are held constant. The direct effect from gender to MC is negative, meaning that there is a decrease for females on MC of -0.94 points for grade three and –1.69 points for grade five when holding all other effects constant. This indicates that when the effect of reading and writing are partialled out, third and fifth grade females are disadvantaged in mathematics for both the MC and CR formats. This finding presents a very different pattern of gender differences than one would conclude by examining only the mathematics scores to the exclusion of verbal abilities.

Gender has a significant direct effect on writing. An average female's writing score is estimated to be 2.81 points higher in grade three, and 3.45 points higher in grade five. However, this is not a reliable prediction because 95.94% of the variability in writing cannot be attributed to gender.

Reading was also hypothesized to be directly influenced by gender. The results of the path analysis provide partial support for this hypothesis. The path from gender to reading is non-significant for grade three, but for grade five the results show that females are advantaged by 0.54 points when all other variables are held constant.

The various paths that involve mediating variables can be combined to show the indirect effects given in Tables 9 and 10 for grades three and five respectively. For example, the indirect effect of gender on MC is made up of the path that leads from gender via writing via reading to MC, and the path that leads from gender via reading to MC. For grade three, the path coefficient from gender to writing is 2.81, from writing to reading is 0.98, and from reading to MC is 0.48. Multiplying these path coefficients results in 1.32, representing the indirect effect of this path. The path that leads from gender to reading is –0.16, and the path from reading to MC is 0.48, which, when multiplied, amounts to –0.08. By adding the indirect effects from gender on MC, one obtains the combined indirect effects 1.32 + (-0.08) = 1.24 (the difference from the exact number 1.25 shown in Table 9 is due to rounding to two decimal places). The other indirect effects given in Tables 9 and 10 are derived in the same manner.

Tables 9 and 10 about here

It is interesting to note that all the significant indirect effects, which involve language abilities (reading and writing), are positive indicating an advantage for females, whereas both direct effects from gender to mathematics MC and CR are negative, indicating an advantage for males when language factors have been partialled out.

The total effects (see Tables 11 and 12), a synthesis of direct and indirect effects, are all positive, which explains why females appear to be performing better than males on mathematics. The effects of writing and reading mask the more subtle gender differences in the performance on mathematics. This masking of gender differences has been overlooked in previous studies. By adopting a simplistic causal model with one single cause (gender) and one single outcome (total mathematics score), the total effects obscure a more complex system in which positive and negative effects balance out or compensate for each other.

Tables 11 and 12 about here

About 58.80% of the variance in the MC score and 68.54% in the CR score are accounted in the causal model for grade three. For grade five, 55.09% of the variance in the MC score and 67.75% in the CR score are accounted for. These proportions are quite substantial and suggest stability of the causal model from grade three to five.

The direct effect of gender on MC does not appear to be equal to the direct effect of gender on CR in either grade three or grade five. To ascertain statistically that the effect of gender on MC is greater than the effect of gender on CR, it was tested whether constraining the parameters for MC and CR to be equal would result in an equally acceptable causal model. The unconstrained model was compared to the constrained model. The chi-square difference tests for grades three and five are given in Table 13. The path coefficients from gender (G) to MC and from gender (G) to CR for the unconstrained model, and the path coefficients from gender (G) to MC/CR in the constrained model, are given in Table 14.

Tables 13 and 14 about here

The results of this statistical procedure for the two grade levels show that the constrained model is inferior to the unconstrained model. This clearly indicates that the direct effect of gender on mathematics performance varies depending on whether the CR or the MC format is involved in the assessment. The path coefficients for grades three and five show that there is a greater disadvantage for females on MC items than on CR items after accounting for language factors.

Discussion

The causal model has been found to be stable from grade three to grade five, with more or less comparable overall fit indices and path coefficients across the two grades. The path coefficients for all paths, except for the path from gender to reading for grade three, are statistically significant. Although total gender difference effect sizes are in favor of females, girls are actually disfavored on both MC and CR after language factors have been accounted for. This finding, hitherto undocumented in Hawaiʻi or on the continental U.S., raises questions as to whether or not the more conventional direct comparison between males and females is the most appropriate or meaningful approach to understanding gender differences in mathematics performance.

The model explains that the reason for the gender difference in favor of females is that the advantage females have in reading and writing improves their mathematics scores. Although males are supposed to have an advantage on mathematics relative to females, males' lower reading and writing scores negatively impact their mathematics performance and mask their relative advantage in this subject. Corroborative evidence from grades eight and ten can be found in Reiss (2005).

The findings of this study have tangible pedagogical implications. Basic literary skills (reading and writing) are pre-requisites to mathematics achievement. For instructional and learning purposes, increasing students' verbal scores might assist in increasing their performance on mathematics assessments. This is especially important for boys, whose lower linguistic skills negatively influence their mathematics assessment.

Because gender differences exist in early literacy skills, mathematics educators may need to consider gender-appropriate pedagogical approaches for boys and girls. To benefit males and females, the instruction for males and females might need to be differentiated. As Gambell and Hunter (2000) stated, "Males are in trouble in literacy!" (p. 712). And as a result, boys are in trouble with mathematics as well. While mathematics performance of males might be improved by focusing on linguistic skills, for females beneficial outcomes might be obtained by focusing on mathematics. Boys might benefit from additional guidance in reading comprehension and verbalization along with quantitative reasoning, whereas for girls the benefit might accrue from focused practice with mathematics-specific semiotics, e.g., symbols, formulas, and algorithms. Because MC and CR items present different challenges to boys and girls, instructors might consider providing girls with opportunities to practice more with MC items, and boys more with CR items.

In view of the consistent finding of girls outperforming boys on SAT, NAEP and HSA, educators in Hawaiʻi need to reconsider the widely adopted assumption of boys being stronger in mathematics. The DOE needs to recognize the need to leave no boys behind in mathematics and language arts.

Limitations

One limitation of this study is that variables such as ethnicity, socioeconomic status, native language, motivation, or parental influences were either unavailable or not accurate enough to be taken into account. Another potential problem is that constructed-response items encompass a wide range of response types ranging from the production of a single word to an essay. It is not clear how different types of CR items may affect boys' and girls' performance differently. This study did not examine students' performance on the various domains of mathematics included in the HSA. It is possible that gender-related performance differences are due to factors not examined in this study, such as cognitive processing requirements as well as linguistic factors.

A further caveat arises from the possibility that CR items, which require more effort to answer than MC items, are skipped more often by males than females. The data set does not identify items on which no attempt was made. Girls may be more conscientious about responding to all items and might have earned more points. Boys, on the other hand, might have given up on CR items on which they might have been able to earn at least some points.

Conclusion

This study built on and extended prior research concerning gender differences in mathematics in Hawaiʻi by providing new understandings about gender differences in mathematics performance. A causal model was confirmed that supports the premise that girls do better than boys in mathematics due to their advantage in reading and writing. After controlling for linguistic factors, girls are found to be disadvantaged on both MC and CR. The disadvantage is more severe on MC than on CR.

Hawaiʻi's unique mathematics test results appear to be due to linguistic factors. While this study provides a plausible explanation for how females and males arrive at their respective CR and MC scores, the reasons and processes accounting for why language factors should affect males in Hawaiʻi more than they might in other places is left for further examination. Future research might consider whether factors such as identity issues and Hawaiʻi Creole English may play a role in the differential performance of gender on mathematics assessment.

The findings of this study call attention to the need for gender appropriate pedagogical approaches to optimize mathematics learning for boys and girls in Hawaiʻi.

Table 1: Sources of Effect Sizes

Study ID	Effect Size No.	Publication	Year Studied	Grade
1	1	(DeMars, 1998)	1996	11
1	2	(DeMars, 1998)	1996	11
2	3	(Zhang & Manon, 2000)	1998	3
2	4	(Zhang & Manon, 2000)	1998	5
2	5	(Zhang & Manon, 2000)	1998	8
2	6	(Zhang & Manon, 2000)	1998	10
2	7	(Zhang & Manon, 2000)	1999	3
2	8	(Zhang & Manon, 2000)	1999	5
2	9	(Zhang & Manon, 2000)	1999	8
2	10	(Zhang & Manon, 2000)	1999	10
3	11	(Wilson & Zhang, 1998)	1995	3
3	12	(Wilson & Zhang, 1998)	1995	5
3	13	(Wilson & Zhang, 1998)	1995	8
3	14	(Wilson & Zhang, 1998)	1995	10
4	15	(Garner & Engelhard, 1999)	1994	11

Table 2: Descriptive Statistics and Effect Sizes for the Multiple-Choice Format

Study ID	Effect Size No.	Males			Females			
		N	M	SD	N	M	SD	g
1	1	572	.606	.196	603	.602	.183	-.02
1	2	652	.607	.211	694	.587	.197	-.10
2	3	3626	32.13	8.96	3463	32.04	8.77	-.01
2	4	3739	32.41	10.09	3777	32.18	9.67	-.02
2	5	3954	25.91	9.94	3681	25.41	9.52	-.05
2	6	3275	22.39	9.5	3276	22.08	8.55	-.03
2	7	3861	34.1	8.78	3674	33.45	8.7	-.07
2	8	4038	32.02	9.79	3790	31.55	9.35	-.05
2	9	3844	26.16	9.71	3719	25.42	9.1	-.08
2	10	3528	22.4	9.42	3407	21.47	8.15	-.11
3	11	4059	19.55	6.17	3854	19.21	5.84	-.06
3	12	3945	20.97	6.79	3789	20.83	6.48	-.02
3	13	3877	22.27	7.52	3807	21.62	6.99	-.09
3	14	3075	17.33	7.29	3129	16.99	6.38	-.05
4	15	1862	43.66	10.4	2090	42.51	10.29	-.11

Table 3: Descriptive Statistics and Effect Sizes for the Constructed-Response Format

Study ID	Effect Size No.	Males			Females			
		N	M	SD	N	M	SD	g
1	1	572	.316	.27	603	.318	.271	.01
1	2	652	.326	.235	694	.343	.213	.08
2	3	4100	7.04	2.67	3871	7.29	2.52	.10
2	4	3948	3.9	2.48	3971	4.13	2.47	.09
2	5	4262	2.93	2.85	3973	2.91	2.88	-.01
2	6	3570	1.57	2.41	3570	1.56	2.34	.00
2	7	4182	6.59	2.88	3915	6.48	2.8	-.04
2	8	4258	2.86	2.81	3985	3.37	2.93	.18
2	9	4148	3.75	3.11	4079	3.8	2.98	.02
2	10	3790	3.11	2.55	3679	3.28	2.33	.07
3	11	4035	16.15	6.31	3871	16.24	6.16	.01
3	12	3889	10.46	6.82	3824	10.07	6.52	-.06
3	13	3974	13.85	7.71	3910	12.57	7.38	-.17
3	14	3095	11.81	6.98	3211	10.88	6.40	-.14

Table 4: Descriptive Statistics for Reading, Writing, Mathematics MC and CR

			Reading		Writing		Math MC		Math CR	
Grade	Sex	N	Mean	STD	Mean	STD	Mean	STD	Mean	STD
3	F	6,352	38.20	10.92	23.46	6.87	26.66	6.93	12.65	6.87
	M	6,354	35.60	11.98	20.65	6.78	26.35	7.41	12.41	7.06
5	F	6,331	36.55	10.53	25.98	6.25	26.57	7.02	14.45	6.96
	M	6,717	32.74	10.81	22.53	6.40	26.33	7.26	13.58	7.21

Table 5: Cronbach’s Alpha for Raw Score Variables

	Math Total	MC	CR	Reading	Writing
Grade 3	0.92	0.88	0.84	0.91	0.96
Grade 5	0.91	0.87	0.83	0.89	0.93

Table 6: Grade Three Variance-Covariance Matrix

	CR	MC	Writing	Gender	Reading
CR	48.61	40.22	24.43	0.06	59.50
MC	40.22	51.54	25.51	0.08	63.27
Writing	24.43	25.51	48.56	0.70	47.52
Gender	0.06	0.08	0.70	0.25	0.65
Reading	59.50	63.27	47.52	0.65	133.06

Table 7: Grade Five Variance-Covariance Matrix

	CR	MC	Writing	Gender	Reading
CR	50.42	40.77	23.34	0.22	56.09
MC	40.77	52.65	22.28	0.06	57.70
Writing	23.34	22.28	42.98	0.86	41.28
Gender	0.22	0.06	0.86	0.25	0.95
Reading	56.09	57.70	41.28	0.95	117.58

Table 8: Goodness of Fit Indices for Grade Three and Five

Grade	Fit Function	GFI	AGFI	CFI	NNI	NFI
3	0.02	0.99	0.90	0.99	0.93	0.99
5	0.01	0.99	0.92	0.99	0.95	0.99

GFI = Goodness of Fit Index
AGFI = Goodness of Fit Index adjusted for degrees of freedom
CFI = Comparative Fit Index
NNI = Non-Normed Index
NFI = Normed Fit Index

Table 9: Indirect Effects for Grade Three

	Gender	MC	Writing	Reading
CR	0.77	0.00	0.42	0.26
MC	1.25	0.00	0.47	0.00
Writing	0.00	0.00	0.00	0.00
Reading	2.75	0.00	0.00	0.00

Table 10: Indirect Effects for Grade Five

	Gender	MC	Writing	Reading
CR	1.17	0.00	0.43	0.27
MC	1.93	0.00	0.48	0.00
Writing	0.00	0.00	0.00	0.00
Reading	3.28	0.00	0.00	0.00

Table 11: Total Effects for Grade Three

	Gender	MC	Writing	Reading
CR	0.24	0.54	0.48	0.43
MC	0.31	0.00	0.47	0.48
Writing	2.81	0.00	0.00	0.00
Reading	2.60	0.00	0.98	0.00

Table 12: Total Effects for Grade Five

	Gender	MC	Writing	Reading
CR	0.86	0.53	0.53	0.45
MC	0.24	0.00	0.48	0.50
Writing	3.45	0.00	0.00	0.00
Reading	3.82	0.00	0.95	0.00

Table13: Comparison of Unconstrained and Constrained Models

	Unconstrained Model		Constrained Model		Difference	
Grade	df	χ^2	df	χ^2	df_{diff}	χ^2_{diff}
3	1	221.82^{**}	2	235.48^{**}	1	13.66^{**}
5	1	172.01^{**}	2	318.43^{**}	1	146.42^{**}

** $p < 0.01$

Table 14: Path Coefficients from Gender to MC and CR

Grade	G to MC	G to CR	G to MC = G to CR
3	-0.94^{**}	-0.54^{**}	-0.71^{**}
5	-1.69^{**}	-0.30^{**}	-0.89^{**}

** $p < 0.01$

Figure 1: Causal Model Depicting Nature of Each Path

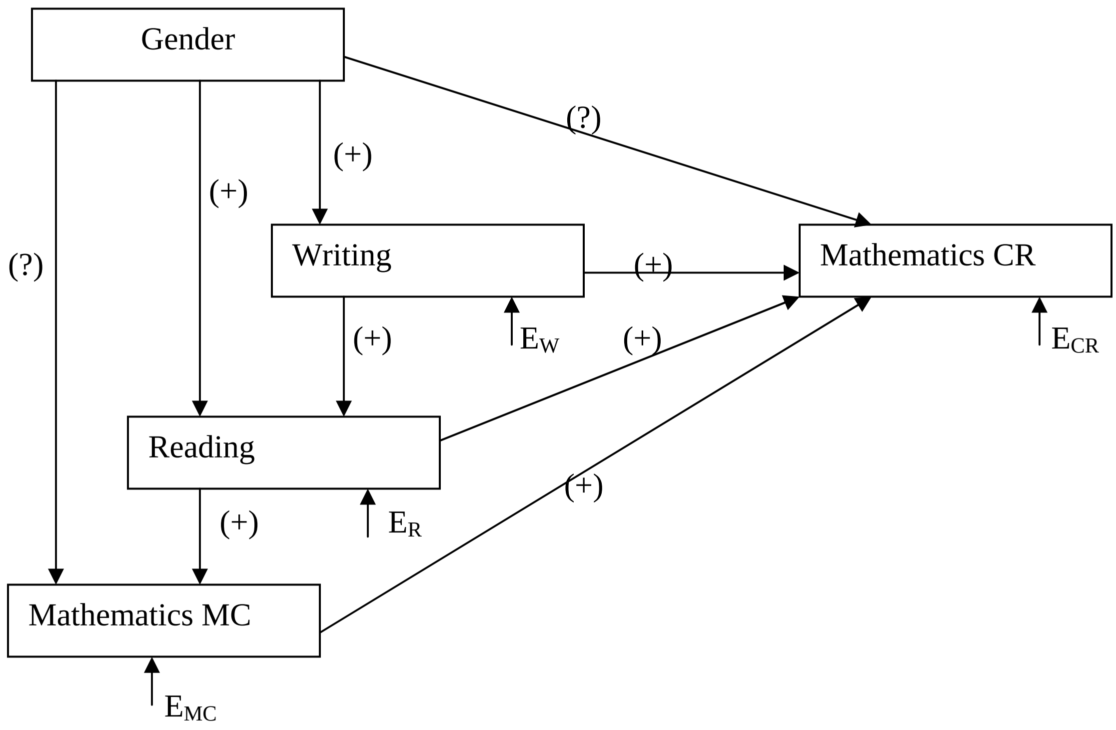
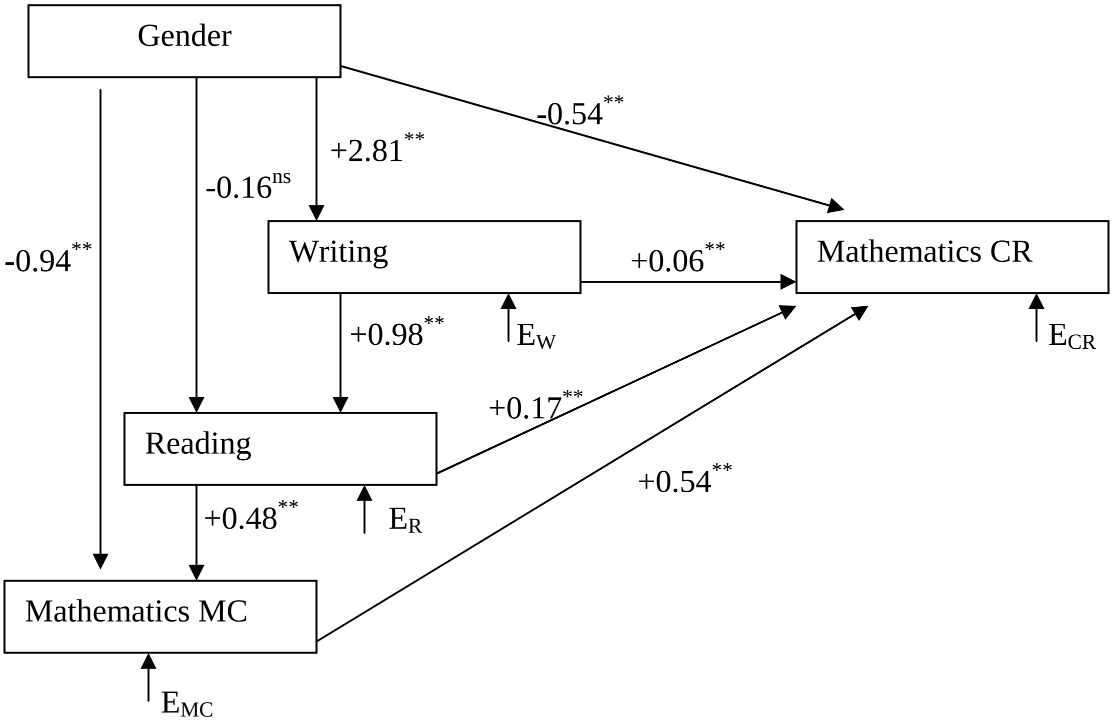


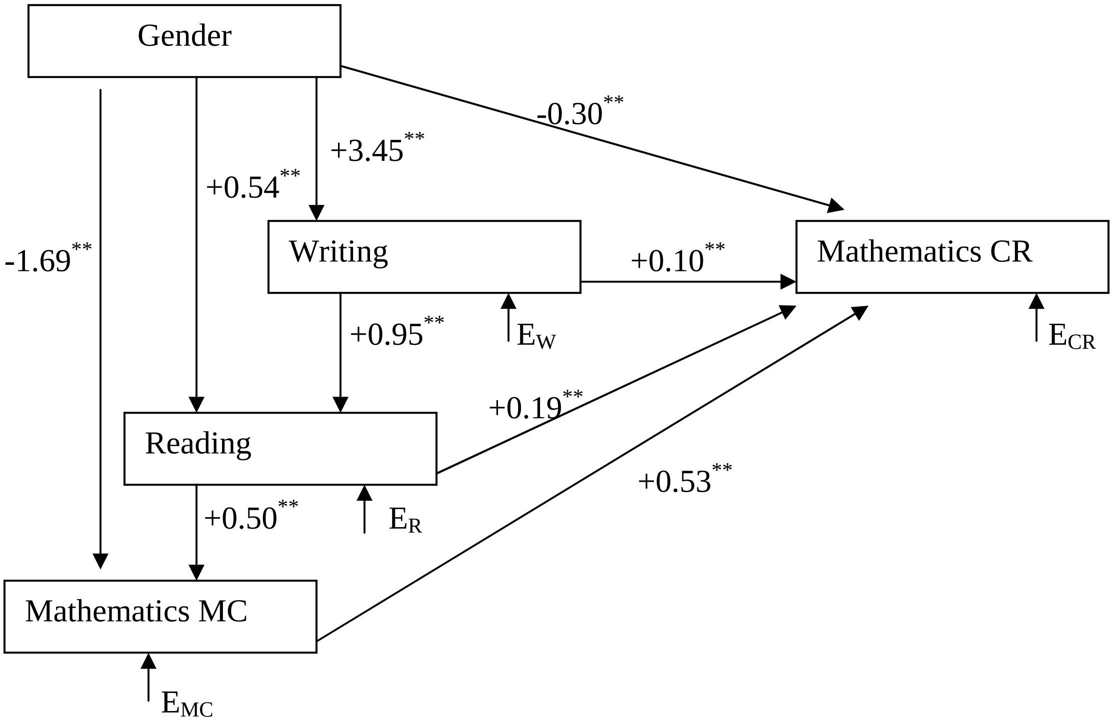
Figure 2: Unstandardized Path Coefficients for Grade Three



** $p < 0.01$

E_W Error Variance in Writing (95.94%)
E_R Error Variance in Reading (65.05%)
E_{MC} Error Variance in Mathematics MC (41.20%)
E_{CR} Error Variance in Mathematics CR (31.46%)

Figure 3: Unstandardized Path Coefficients for Grade Five



$^{**}p < 0.01$

E_W Error Variance in Writing (93.08%)
E_R Error Variance in Reading (66.22%)
E_{MC} Error Variance in Mathematics MC (44.91%)
E_{CR} Error Variance in Mathematics CR (32.33%)

References

Abedi, J. (2000). *Confounding of students' performance and their language background variables* (ERIC Document Reproduction Service No. ED 449250): Urban Education.

Abedi, J., Lord, C., & Plummer, J. (1995). *Language background as a variable in NAEP performance: NAEP task 3D: Language background study (CSE Tech. Rep No. 429)*. Los Angeles: UCLA Center for the Study of Evaluation/National Center for Research on Evaluation, Standards, and Student Testing.

Brandon, P. R., & Jordan, C. (1994). Gender differences favoring Hawai'i girls in mathematics achievement: Recent findings and hypotheses. *Zentralblatt fuer Didaktik der Mathematik, 94*(1), 18-21.

Brandon, P. R., Newton, B. J., & Hammond, O. W. (1987). Children's mathematics achievement in Hawaii: Sex differences favoring girls. *American Educational Research Journal, 24*(3), 437-461.

Cole, N. S. (1997). *The ETS gender study: How females and males perform in educational settings* (No. ED 424337). Princeton, NJ: Educational Testing Service.

Coley, R. J. (2001). *Differences in the gender gap: Comparisons across racial/ethnic groups in education and work*. Princeton: Educational Testing Service (ED 451 222).

De Avila, E. A. (1988). Bilingualism, cognitive function, and language minority group membership. In R. R. Cocking & J. P. Mestre (Eds.), *Linguistic and cultural influences on learning mathematics* (pp. 101-121). Hillsdale: Erlbaum.

DeMars, C. E. (1997). *Physics or biology? Geometry or algebra? Gender and content area interactions on a high school proficiency test.* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago.

DeMars, C. E. (1998). Gender differences in mathematics and science on a high school proficiency exam: The role of response format. *Applied Measurement in Education, 11*(3), 279-299.

DeMars, C. E. (2000). Test stakes and item format interactions. *Applied Measurement in Education, 13*(1), 55-77.

Gambell, T., & Hunter, D. (2000). Surveying gender differences in Canadian school literacy. *Journal of Curriculum Studies, 32*(5), 689-719.

Garner, M., & Engelhard, G. J. (1999). Gender differences in performance on multiple-choice and constructed response mathematics items. *Applied Measurement in Education, 12*(1), 29-51.

Gronna, S., Chin-Chance, S., & Abedi, J. (2000, April). *Differences between the performance of limited English proficient students and students who are labeled proficient in English on different content areas: reading and mathematics.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans.

Halpern, D. F. (2000). *Sex differences in cognitive abilities* (3rd ed.). Mahwah, NJ: Erlbaum.

Halpern, D. F. (2004). A cognitive-process taxonomy for sex differences in cognitive abilities. *Current directions in psychological science, 13*(4), 135-139.

Hawai‘i State Department of Education. (2003). *NAEP 2003 math -- Hawai`i keeps pace with national gains*. Retrieved June 25, 2005, from http://lilinote.k12.hi.us/STATE/COMM/DOEPRESS.NSF

Hawai‘i State Department of Education. (2001). *NAEP: Hawai`i's 4th- and 8th-grade math results*. Retrieved June 25, 2005, from http://lilinote.k12.hi.us/State/Comm/DOEPRESS.NSF/0/1231b444216bb45b0a256a9c

Hawai‘i State Department of Education. (2005). *Accountability Resource Center Hawai‘i*. Retrieved February 04, 2005, from http://arch.k12.hi.us/school/default.html

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin, 104*, 53-69.

Kiplinger, V. L., Haug, C. A., & Abedi, J. (2000, April). *Measuring math -- not reading -- on a math assessment: A language accommodations study of English language learners and other special populations.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans.

Maccoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex differences*. Stanford, CA: Stanford University Press.

Myerberg, N. J. (1996). *Performance on different test types by racial/ethnic group and gender.* Paper presented at the Annual Meeting of the American Educational Research Association, New York.

Nowell, A., & Hedges, L. V. (1998). Trends in gender differences in academic achievement from 1960 to 1994: An analysis of differences in mean, variance, and extreme scores. *Sex Roles, 39*(1/2), 21-43.

Office of Educational Accountability. (2002). The Minnesota basic skills test: Performance gaps for 1996 to 2001 on the reading and mathematics tests, by gender, ethnicity, limited English proficiency, individual education plans, and socio-economic status: Minnesota University, Minneapolis. Office of Educational Accountability. (ERIC Document Reproduction Service No. ED 470 555).

Reiss, P. P. (2005). *Causal models of item format- and gender-related differences in performance on a large-scale mathematics assessment for grade three to grade ten.* Unpublished doctoral dissertation, University of Hawai`i at Manoa, Honolulu.

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (revised ed. Vol. 6). Thousand Oaks, CA: Sage.

Willingham, W. W., & Cole, N. S. (1997). *Gender and fair assessment*. Mahwah, NJ: Erlbaum.

Wilson, L. D., & Zhang, L. (1998, April). *A cognitive analysis of gender differences on constructed-response and multiple-choice assessments in mathematics.* Paper presented at the Annual Meeting of the American Educational Research Association, San Diego.

Zhang, L., & Manon, J. (2000). *Gender and achievement - understanding gender differences and similarities in mathematics assessment*. Retrieved July 24, 2004, from http://www.doe.state.de.us/aab/Gender%20Research.pdf