科技部補助專題研究計畫成果報告
期末報告

以TIMSS 2011台灣學生科學成就資料檢視試題類型、內容領域與認知階層中的性別差異(A06)

計 畫 類 別 ： 個別型計畫
計 畫 編 號 ： MOST 104-2629-S-008-001-
執 行 期 間 ： 104年08月01日至105年07月31日
執 行 單 位 ： 國立中央大學學習與教學研究所

計畫主持人： 劉佩艷

計畫參與人員： 碩士班研究生-兼任助理人員：王呈隆
碩士班研究生-兼任助理人員：黃思婧
碩士班研究生-兼任助理人員：梁長雯
大專生-兼任助理人員：陳嘉恩

報 告 附 件 ： 出席國際學術會議心得報告

中 華 民 國 105 年 10 月 31 日

中 文 摘 要 ： 本研究旨在檢驗四年級與八年級學生在性別與兩測驗構成相關因子，即試題類型與認知領域間交互影響下之科學成就表現。台灣資料的部分，係取自於國際數學與科學教育成就趨勢調查（Trends in International Mathematics and Science Study，簡稱TIMSS）2011年之施測報告，並使用多向度Rasch模式估量學生各個部分能力。研究結果顯示，女生在選擇題中表現顯著較男生來得差，不過在建構反應題中表現較佳；同時，由四年級至八年級，試題類型對學生科學成就之影響增加。此外，儘管四年級女生的認知能力就整體以及認識領域而言，顯著低於男生，但其在推理領域的認知能力顯著高於男生，而八年級學生中，則沒有發現整體及各認知領域中有性別差異。本研究結果揭示試題類型對學生科學成就在性別間可能存在不同作用效果。

中 文 關 鍵 詞 ： 性別差異、科學成就、國際數學與科學教育成就趨勢調查、試題類型、認知領域

英 文 摘 要 ： The purpose of this study was to examine fourth and eighth grade students’ science performance in terms of the interaction between gender and two test-related components, item format and cognitive domains. The portion of Taiwanese data came from the 2011 administration of the Trends in International Mathematics and Science Study (TIMSS). A multidimensional Rasch model was used to estimate student abilities for each component. The results indicated that girls performed significantly worse than boys in multiple-choice items, but did better in constructed-response items. The impact of item format on student science achievement enlarges from the fourth to the eighth grade. While fourth grade girls’ cognitive ability in general and in knowing is significantly lower than boys’, girls’ cognitive ability in reasoning is statistically higher than boys. No gender difference was found in the cognitive domain in general and different domains in the eighth grade. The results of this study shed light the possible differential function of the item format on student science achievement by gender.

英 文 關 鍵 詞 ： cognitive domain, gender difference, item format, science achievement, TIMSS

Gender Differences in Item Formats, Content Domains, and Cognitive Domains: Results from TIMSS 2011 Taiwanese Student Science Data

Abstract

The purpose of this study was to examine fourth and eighth grade students' science performance in terms of the interaction between gender and two test-related components, item format and cognitive domains. The portion of Taiwanese data came from the 2011 administration of the Trends in International Mathematics and Science Study (TIMSS). A multidimensional Rasch model was used to estimate student abilities for each component. The results indicated that girls performed significantly worse than boys in multiple-choice items, but did better in constructed-response items. The impact of item format on student science achievement enlarges from the fourth to the eighth grade. While fourth grade girls' cognitive ability in general and in knowing is significantly lower than boys', girls' cognitive ability in reasoning is statistically higher than boys. No gender difference was found in the cognitive domain in general and different domains in the eighth grade. The results of this study shed light the possible differential function of the item format on student science achievement by gender.

Keywords: cognitive domain, gender difference, item format, science achievement, TIMSS

Introduction

Differential gender performance in science has been a concern shared among policy makers, education researchers, educators, and stakeholders. The gender differences shape students' self-concept of learning science by gender, which in turn affects their achievements in science (Marsh & Yeung, 1997; Marsh & O'Mara, 2008). Science achievement and self-concept of learning science are two of the key determinants for students to choose a college major, and further, a career in the fields of science, technology, engineering and mathematics (STEM), which is viewed as a predictor for both individual economic status and national power. However, the literature indicates that there is gap between males and females in such fields (Mearten-Rivera, Myers, Lee, & Penfield, 2010; National Science Foundation, 2013; Reis & Park, 2001). A paucity of females in the fields of STEM reduces gender diversity, and probably also reduces the diverse knowledge and experiences which inform both societal policy-making and implementation.

Many studies from different perspectives have been conducted to broaden our understanding the gender differences over the decades (e.g., Amelink, 2009; Becker, 1989; Lee & Burkam, 1998). Biological, sociological, and psychological elements account for differences in the participation of underrepresented groups in STEM learning and workforce have been proposed. Supports for, or barriers to these physical, social, cultural, and cognitive participation in STEM learning and workforce have been executed. However, these elements are not malleable. On the other hand, the formats of assessments for measuring students' science knowledge in school settings can be relatively malleable. Therefore, it is worthy for science education researchers, educators, and practitioners to investigate how the components of science assessment influence the possible gender differences of science achievement.

Student science achievement is often referred to as a homogeneous single global construct. However, the composition of its whole score is rarely paid attention to not only in terms of how it is measured but also what it is meant to measured. In the setting of educational measurement, the impact of item formats on student performance has been an issue (e.g., DeMars, 1998; Kan & Bulut, 2014; Liu & Wilson, 2009a; Liu & Wilson, 2009b). Item formats may have differential functioning in achievement by gender. For instance, Penner (2003) showed that males perform better on multiple choice and females work better on open-ended items. Meanwhile, science achievement is often measured using a test which is often composed of distinctive cognitive domains, such as knowing, applying, and reasoning (Martin, Mullis, Foy, & Stanco, 2012). While there may be a gender difference in overall science achievement, the patterns across domains may not be consistent. Therefore, there is a need to investigate the possible differential functions of item formats in student science achievement across cognitive domains.

Therefore, this study intends to examine the effects of item format and cognitive domains on science performance by gender through a more comprehensive lens. Gender differences may be underestimated by comparing performance based on the total test score. If boys and girls are each favored by certain item formats and cognitive domains, then the effect of gender differences may get cancelled out when a total score is used. This disaggregation would prevent making invalid inferences about student achievement in science. Therefore, this study employed a multidimensional analysis approach to distinguish science achievement depending on item format and cognitive domains. The results of this study shed light on the patterns and particular gender differences to identify the two factors that are likely to introduce statistical differences. The results of this study could serve as guidelines for the development of achievement testing when designing gender-based initiatives to address gender differences in science achievement.

Item Formats

While examining the gender differences in science scores from the student side is important, it is even more important to engage in discourse on the gender gap from the assessment side, which can be manipulated by educators, practitioners, and test developers. Amelink (2009) highlighted that assessment methods, including item formats, have an influence on the magnitude of gender difference in science performance. Some research findings (DeMars, 1998; Harding, 1979; Hoste, 1982; Murphy, 1982) have found that males scored higher on multiple choice items, whereas females are favored by constructed response items. As multiple choice items normally constitute a large portion of science assessment, males may earn higher total scores. Multiple choice items are particularly favorable in large-scale assessments since they are low cost compared to alternative formats (e.g., open ended items) (Lawrenz, Huffman, & Welch, 2000).

Derived from ILSA data, several studies, using the Programme for International Student Assessment (PISA) data, have examined the association between item format and student performance by gender. Liu and Wilson (2009a) analyzed the PISA 2000 and 2003 U.S. mathematics data, and their results indicated that males have significant advantages on complex multiple-choice items over females. Liu and Wilson (2009b) further extended the idea to examine PISA 2003 Hong Kong mathematics data, the results of which were also consistent with those from the U.S. By analyzing PISA 2003 Hong Kong student data, Yip, Chiu and Ho (2004) concluded that boys scored higher on closed items. While these studies (e.g., Liu & Wilson, 2009a, 2009b) have investigated the gender differences due to item formats in large-scale assessment mathematics data, there is relatively little peer-reviewed literature on large-scale assessment science data. Yip, Chiu and Ho's study (2004) may be the only case; however, the results are from PISA student science data, not TIMSS data. Compared to PISA, the items in TIMSS are more school curricula-based. The results from TIMSS may thus have greater implications for science instruction and assessment.

Cognitive Domains

When interpreting total scores, the distinct cognitive domains consisting of the test are also often neglected. In the cognitive domains, the nature of science achievement is heterogeneous. Bloom, Engelhart, Furst, Hill and Krathwohl (1956) proposed a taxonomy of educational objectives to shed light on the hierarchical nature of cognitive domains. These domains include knowledge, comprehension, application, analysis, synthesis, and evaluation, each of which depends on the preceding one. Higher-level cognitive skills are considered to be more valuable than the knowledge of simple and discrete facts, since they involve the ability to solve more complex tasks. A student able to solve problems in a higher-level cognitive domain is viewed as being equipped with a more sophisticated ability than mere surface rote memorization. Owning higher-level cognitive domain is an essential skill for success in life and STEM careers. Thus, it may be assumed that items composed of hierarchical levels of cognitive domains, such as knowing, applying and reasoning, may have different impacts on students' performance.

Concern about the issue of equality in science assessment and education has been raised. However, to date, no related study that examined the impact of item format and cognitive domains on student science achievement by gender has been found. Thus, this study intends to fill in this literature gap.

Research Questions

While the gender difference is contextualized and varies across different science tests, this study seeks to provide a better understanding of the role of item format and cognitive domain on Taiwanese fourth and eighth grade student science achievement based on national representative data, TIMSS 2011. In each cycle of TIMSS 2011, while some of the items are not available for public use, others are released to the public. By utilizing these released items, the following research questions can be pursued.

1. To what extent is the gender difference in overall science scores attributable to the format of the items in the fourth and eighth grades, respectively?
2. To what extent is the gender difference in overall science scores attributable to the cognitive domains of the items in the fourth and eighth grades, respectively?

3. To what extent is the magnitude of gender difference in student science scores due to item formats and cognitive domains from the fourth to the eighth grades?

Methods and Analyses

Data Source and Samples

The data used in this study came from the Taiwanese portion of the TIMSS 2011 data. TIMSS 2011 are the most up-to-date standardized test data released for public use. TIMSS is designed to provide cross-national information about eight- and fourth-grade student science and mathematics achievement and related background variables. Two-stage stratified sampling was adapted to survey students representing the target population. First, the schools were selected based on the probability proportional to the school size (PPS) after stratification. One of the intact classrooms within the selected schools was randomly chosen. All students in the chosen classrooms were further surveyed.

TIMSS intends to measure students' science scores in terms of a broad range of science curricula. In order to achieve the goal, 172 items in the fourth grade and 217 items in the eighth grade are designed to elicit students' science proficiency. However, it would be impractical and time-consuming for all students to take all of the items. Thus, a complex matrix-sampling booklet design was utilized. The matrix-sampling booklet design packs the entire assessment pool of science items to a set of 14 student achievement booklets, with each student answering just one booklet. Plausible values were randomly drawn from the distribution of ability estimates that represent the range of reasonable values for a students' ability. Details of the data, tests and sampling procedures of TIMSS 2011 can be found in the technical reports (Martin et al., 2012; Mullis, Martin, Ruddock, O'Sullivan, & Preuschoff, 2009).

A total of 2,444 fourth and 2,886 eighth grade students in Taiwan were the samples. The fourth grade students' data including 1,166 girls (47.7%) and 1,278 boys (52.3%) as well as eighth grade students' data, including 1,402 girls (48.58%) and 1,484 boys (51.42%) were investigated in this study. In the original TIMSS 2011 Taiwanese data, 4,284 fourth and 5,042 eighth graders were surveyed. The percentage for fourth grade girls is 46.90% (N=2029) and boys is 53.10% (N=2242) as well as for the eighth grade girls and for boys is boys is 51.57% (N=2,594). `Due to the complex matrix-sampling booklet and only partial exam items utilized in this study, there is too much missing data for some of the students. Thus, only 2,444 fourth and 2,886 eighth grade students were selected as samples to be examined.

Measures

Four variables were utilized in this study, namely 1) item format, 2) cognitive domains, 3) students' science achievement scores by specific categories created by the researchers, and 4) gender. Due to the scaling for creating equivalent scores over years, only a portion of the total items in TIMSS 2011 was revealed completely for public use. The remaining items are kept confidential for use in the future cycles. Thus, the present study focuses solely on the release items to allow for examination of the specific item format and cognitive domains. The exact item formats of the 72 items for the fourth grade and 88 items for the eighth grade can be examined and determined by three coders. Two of the coders are research assistant whose backgrounds are in science education and chemistry. The third code is the first author whose academic background is measurement and applied statistics in the field of education research and has a publication record in science education. The following sections regarding the first three variables were stated.

Item format: Of the released 72 fourth grade items, their format was categorized into either multiple-choice or constructed-response items by the researchers. 61% of the items were the multiple-choice and the other 39% are the constructed response items. The multiple-choice items refer to several options being available for the students to choose from, where only one is the best possible answer. There are four options for each multiple choice item. The types of constructed-response items include short response, closed constructed-response, and open constructed-response items. Specifically, the short response items are a form in which students are asked to write words provided in the item contents. The definition of the closed constructed response items is that the correct answer to the question is provided and the students need to write their arguments for it. On the other hand, for the open constructed response items, the students have to write a paragraph to support their argument; the items may not necessarily have an absolute yes/no answer. On the other hand, at the eighth grade, the released 88 items include 44 items (50%) were multiple-choice and the other half items (50%) were constructed response items.

Cognitive domain: At the fourth grade level, 30 items (41.67%) were located in the domain of knowing, 31 items (42.06%) in applying, and 11 items (15.28%) in reasoning. At the eighth grade level, 32 items (36.36%) were located in the domain of knowing, 34 items (38.63%) in applying, and 22 items (25%) in reasoning. Compared to the original design of TIMSS science assessment (Mullis et al., 2009), which was composed of 40% and 35% knowing, 40% and 35% applying as well as and 20 % and 30% reasoning in the fourth and eighth grade, respectively, the percentages of the released items in the reasoning domain seemed to be smaller. The two-way cross table of 72 items at the fourth grade and 88 items at the eighth grade formed by item formats and cognitive domain is listed in Tables 1 and 2.

Table 1. Distribution of 72 Science Items by Item Format and Cognitive Domains for the Fourth Graders

Items	Knowing	Applying	Reasoning	Total
Multiple-choice	20	18	6	44
Constructed response	10	13	5	28
Total	30	31	11	72

Table 2. Distribution of 88 Science Items by Item Format and Cognitive Domains in the Eighth Grade

Items	Knowing	Applying	Reasoning	Total
Multiple-choice	22	15	7	44
Constructed response	10	19	15	44
Total	32	34	22	88

Student science achievement scores in each category: Not every student answered all of the released 72 and 88 items due to the booklet design. The students merely completed the items presented in their own assigned booklet. Due to the time constraints, and given the intention of measuring students' broad knowledge, students were administered only a subset of the items from a larger pool (Martin & Mullis, 2012). For science assessment, the items are separated into 14 booklets. Students are only assigned one of these booklets. Each item appears in two booklets, and it provides a mechanism for linking together the student responses from the 14 booklets. Since the target items in this study are the released 72 and 88 items, students' subscores in each item format and cognitive domain were created based on a multidimensional Rasch model, which is introduced shortly in the subsection on the statistical analyses.

Statistical Analyses

Two quantitative techniques were utilized to examine the gender differences in science achievement across items formats and cognitive domains. First, the Rasch subdimension model (Brandt, 2008) was applied to compute students' science scores in each category. The model is a special case of a bi-factor Rasch model by incorporating an additional set of parameters for estimating an overall ability and subscores for each subdimension. For instance, the model provided an overall estimate of ability as well as two subscores (one for multiple-choice items and another for constructed-response items) for the item format component. Thus, not only the students' general science ability but also specific abilities within the general ability were estimated in the same model. Therefore, the model allows determining the strengths and weakness in each subdimension. The maximum a posteriori (MAP) estimation was utilized. The mean of the estimated scores is zero, and the standard deviation is one. The software program *ConQuest* (Wu, Adams, & Wilson, 2007) was used to perform the subscore estimation.

Furthermore, independent-sample t tests were implemented to determine whether the magnitude of the gender differences or similarities was significant. Independent-sample t tests were used to indicate the ratio of the estimated mean gender difference and the estimated standard error of the mean. The null hypothesis is that there is no difference between girls' and boys' scores (i.e., mean difference equals zero), and the t statistics is considered significant at the alpha equal 0.05 level.

Results

Tables 2 and 3 presents the descriptive statistics and gender comparison based on each specific domain for the fourth and eighth grade students, respectively. The numbers under "mean" represent the science scores estimates produced by *ConQuest*. The larger the value, the higher the average science score for that specific domain. The mean performance difference between girls and boys is also provided, along with the estimated standard error of the mean difference. *t* statistic is used to indicate the significance of the gender differences, and a positive *t* value suggests a girl advantage.

Results of the Fourth Graders

Overall, in the item format domain, fourth grade students' had better scores in multiple-choice (M = .073035) than constructed-response items (M = -.157274). The impact of overall item format on gender difference is statistically significant (t = -4.938, p < .01) where boys (M = .395217) were more favored than girls (M = .291272). When comparing the means of each item format type by gender, boys (M = .073035) had a significantly higher advantage (t = -4.242, p < .01) than girls (M = .063183) in the multiple-choice items. On the contrary, girls (M = -.136130) obtained significantly higher scores (t = 4.229, p < .01) than boys (M = -.176564) in the construct-response items.

In terms of the cognitive domain, students obtained highest scores in knowing (M = .000860), followed by applying (M = .000285) and reasoning (M = -.001145). Boys (M = .481338) obtained significantly higher scores (t = -4.825, p < .01) than girls (M = .357285) in the overall cognitive domain. Boys (M=.001409) had a significantly higher advantage (t = -4.055, p < .01) than girls (M = .000258) in knowing. In applying, while no statistical significance exists, girls (M = .000548) had higher scores than boys (M = .000045). Furthermore, girls (M = -.000806) had a significant advantage (t = 2.096, p < .05) than boys (M = -.001455) in reasoning. The estimated scores and statistical testing of each category of the fourth graders by gender in Table 3.

[insert Table 3 around here]

Results of the Eighth Graders

Generally, the eighth grade students had an advantage in multiple-choice (M =.003556) than the constructed-response items (M = -.003563). There is no statistically significance (t = -1.425) between boys (M = .168248) and girls (M = .126920) in the overall item format domains. Specifically, boys (M =.007245) had significantly higher scores (t = -5.381, p < .01) than girls (M = -.000349) in multiple-choice items. When answering the construct-response items, girls (M =.000238) had a significant advantage (t = 5.341, p < .01) than boys (M = -.007154).

In the cognitive domains, students had better scores in knowing (M = .000110), followed by applying (M = .000038) and reasoning (M = -.000148). Boys (M = .176808) had higher scores than girls (M = .132959) in the overall cognitive domains, but no significant difference existed (t = -1.496). When comparing the means by gender in each cognitive domain, no significant differences were found. In knowing, boys (M =.000315) had higher scores than girls (M = -.000107). However, in the higher cognitive levels, girls had higher scores (M =.000166, -.000059) than boys (M = -.000083, -.000232) in applying and reasoning. The estimated scores and t-test of each domain by gender of the eighth graders in Table 4.

[insert Table 4 around here]

Comparisons of the Results of Fourth and Eighth Graders

The results indicated that girls had a significant disadvantage than boys in multiple-choice items, but had a significant advantage in constructed-response items in both grades. The gap of gender differences seems to be wider in the eighth grade than the fourth grade in individual item formats. However, it is also worth to noticing that the overall item formats, combing the effects of multiple choice and constructed response, on gender differences did not differ. On the other hand, in terms of the cognitive domain, boys performed better in lower-level cognitive domains than girls. That is, girls perform better in reasoning rather than in knowing. The gap of gender differences seems to be narrower in the eighth grade than in the fourth grade. In sum, due to a higher percentage of items on the constructed response and reasoning domains on the eighth grade exam, no significant gender difference is found.

Conclusions and Implications

This study aims to investigate the patterns of gender difference by item formats, and cognitive domains, which are likely to introduce "unfair" student science achievement. While gender-equitable assessments are an ideal for educators, practitioners, and stakeholders to pursue, there should be more valid research studies exploring potential factors for making equitable assessments of the achievement of girls and boys in science learning. The findings of research studies serve such a purpose. Not only does it address the issue of equality, but this study also adds to the body of work on the issues of measurement and TIMSS data analysis in the field of science education. The current investigation specifically aims at answering whether there is gender differences on (1) the item formats employed to elicit students' responses, and (2) the hierarchical cognitive domains measured by the achievement test assessment based on the TIMSS 2011 Taiwanese eighth grade students' data.

This study aims to make two major contributions to the field of science education research. First, the significance of this study lies in revealing the influences of items formats and cognitive domains on student science achievement. The nature of student science achievement, composed of different cognitive domains and being measured by different item formats, is heterogeneous. Even though the issue of equality in science assessment is both practically and academically important, few studies have been conducted to examine such relationships between the two factors and science achievement by gender. Thus, the findings of this investigation contribute to a better understanding of the role of item formats and cognitive domains on gender differences. From the implications of the findings, educators and test developers should be informed of the item formats and distinctive cognitive domains that have been identified as potential sources of gender bias.

Second, the uniqueness of this study is its examination of the gender-equitable issue in the setting of science standardized achievement tests using the Taiwanese portion of the internationally well-recognized TIMSS 2011 data. While many studies have utilized ILSA to investigate various educational phenomena, there seems to be no studies devoted to this measurement issue in science education, either internationally or locally. This study focuses on data from Taiwan to provide evidence-based arguments to inform local practice. The implications of this study may be utilized by test developers, teachers, practitioners and stakeholders in Taiwan with valid results for reference. Moreover, developing gender-equitable assessment is an international topic. Thus, it can be anticipated that the results of this study can contribute to the literature and benefit science educational practices, both domestically and internationally.

Two issues need to be discussed to strengthen the validity of this study. First, only students' responses to science items, but not other contextual variables, were used to estimate students' science scores in each domain of item formats and cognitive levels. In the public released TIMSS data, students' science achievement were imputed based on only not students' responses to exam items, but also other detailed student background information. The focus of TIMSS is to depict the general student science achievement and characteristics of the target population, the whole eighth graders in the country, but not individual students. Additionally, due to the booklet design, too much missing value for students' responses to items. Thus, the exact scores for individual student were not computed in the original TIMSS data. Instead, five plausible values were drawn for constructing the population distributions of students' science achievement. However, the approach of this study is to estimate individuals' science scores instead of imputing plausible values. Thus, the inferential statistics of the two approaches for the intended populations are not identical. Second, the dichotomous categorization of item formats (i.e., multiple-choice and constructed response items) may be too simplified to well represent the formats of items. In the current research design, very limited released items and a large portion of student item responses are missing, so further complex categorization of item formats is not plausible. It is strongly encouraged that more categorization of item formats should be quantified in the future studies.

References

Amelink, C. (2009). *Literature overview: Gender differences in science achievement.* Retrieved January 10, 2015, from http://www.engr.psu.edu/awe/misc/arps/arp_genderdifferencesscience_overview.pdf

Becker, B. J. (1989). Gender and science achievement: A reanalysis of studies from two meta-analyses. *Journal of Research in Science Teaching, 26*(2), 141-169.

Bloom, B. S., Englehart, M. B., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives. The classification of educational goals. Handbook I: Cognitive domain*. New York: Longmans Green.

Brandt, S. (2008). Estimation of a Rasch model including subdimensions. In M. von Davier & D. Hastedt (Eds.), *IERI monograph series: Issues and methodologies in large-scale assessments* (Vol. 1, pp.51-70). Princeton, NJ: IEA-ETS Research Institute.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

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.

Harding, J. (1979). Sex differences in examination performance at 16+. *Physics Education, 14*(5), 280-284.

Hoste, R. (1982). Sex difference and similarities on performance in a CSE biology examination. *Educational Studies, 8*(2), 141-153.

Kan, A., & Bulut, O. (2014). Examining the relationship between gender DIF and language complexity in science assessment. *International Journal of Testing, 14*(3), 245-264.

Lawrenz, F., Huffman, D., & Welch, W. (2000). Policy considerations based on a cost analysis of alternative test formats in large scale science assessments. *Journal of Research in Science Teaching, 37*(6), 615-626.

Lee, V. E., & Burkam, D. T. (1998). Gender differences in middle grade science achievement: Subject domain, ability level, ad course emphasis. *Journal of Research in Science Teaching, 80*(6), 613-650.

Liu, O. L., & Wilson, M. (2009a). Gender differences in large-scale mathematics assessments: PISA trend 2000 & 2003. *Applied Measurement in Education, 22*(2), 164-184.

Liu, O. L., & Wilson, M. (2009b). Gender differences and similarities in PISA 2003 mathematics: A comparison between the United States and Hong Kong. *International Journal of Testing, 9*(1), 20-40.

Marsh, H. W., & O'Mara, A. (2008). Reciprocal effects between academic self-concept, self-esteem, achievement, and attainment over seven adolescent years: Unidimensional and multidimensional perspectives of self-concept. *Personality and Social Psychology Bulletin, 34*(4), 542-552.

Marsh, H. W., & Yeung, A. S. (1997). Causal effects of academic self-concept on academic achievement: Structural equation models of longitudinal data. *Journal of Educational Psychology, 89*(1), 41-54.

Martin, M. O., & Mullis, I. V. S. (Eds.). (2012). *Methods and procedures in TIMSS and PIRLS 2011*. Chestnut Hill, MA: Boston College.

Martin, M. O., Mullis, I. V. S., Foy, P., & Stanco, G. M. (2012). *TIMSS 2011 international results in science.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mearten-Rivera, J., Myers, N., Lee, O., & Penfield, R. (2010). Student and school predictors of high-stakes assessment in science. *Science Education, 94*(6), 937-962.

Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., & Preuschoff, C. (2009). *TIMSS 2011 assessment frameworks.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Murphy, R. J. L. (1982). Sex differences in objective test performance. *British Journal of Educational Psychology, 52*(2), 213-219.

National Science Foundation. (2013). *Women, minorities, and persons with disabilities in science and engineering: 2013* (Special Report NSF 13-304). Arlington, VA. Retrieved from http://www.nsf.gov/statistics/wmpd/

Penner, A. M. (2003). International gender × item difficulty interactions in mathematics and science achievement tests. *Journal of Educational Psychology, 95*(3), 650-655.

Reis, S. M. & Park, S. (2001). Gender differences in high-achieving students in math and science. *Journal of the Education of the Gifted, 25*(1), 52-73.

Wu, M. L., Adams, R. J., & Wilson, M. R. (2007). ConQuest version 2.0: Generalized item response modeling software [Computer software]. Melbourne, Australia: ACER

Yip, D. Y., Chiu, M. M., & Ho, E. S. C. (2004). Hong Kong student achievement in OECD-PISA study: Gender differences in science content, literacy skills, and test item formats. *International Journal of Science and Mathematics Education*, *2*(1), 91-106.

Table

Table 3. *The Estimated Scores and Statistical Testing of Each Category of the Fourth Graders by Gender*

	Total		Girls		Boys			
	Mean (s.e.)	Std. Dev	Mean (s.e.)	Std. Dev	Mean (s.e.)	Std. Dev	t-value	d
Overall item format ability	.345626 (.0105853)	.5233058	.291272 (.0148944)	.5085936	.395217 (.0148744)	.5317468	-4.938**	-0.200
Multiple-choice	.073035 (.0022305)	.1102704	.063183 (.0031431)	.1073255	.082023 (.0031380)	.1121795	-4.242**	-0.171
Constructed-response	-.157274 (.0048020)	.2373940	-.136130 (.0067685)	.2311237	-.176564 (.0067541)	.2414528	4.229**	0.171
Overall cognitive ability	.422154 (.0129248)	.6389602	.357285 (.0182055)	.6216579	.481338 (.181530)	.6489536	-4.825**	-0.195
Knowing	.000860 (.0001422)	.0070297	.000258 (.0002052)	.0070076	.001409 (.0001960)	.0070076	-4.055**	-0.164
Applying	.000285 (.0001351)	.0066783	.000548 (.0001896)	.0064734	.000045 (.0001917)	.0068537	1.865	0.075
Reasoning	-.001145 (.0001543)	.0076299	-.000806 (.0002270)	.0077499	-.001455 (.0002100)	.0075085	2.096*	0.085

Note. ** indicates the value is significant at 0.01 level. * indicates the value is significant at 0.05 level.

Table 4. *The Estimated Scores and Statistical Testing of Each Category of the Eighth Graders by Gender*

	Total		Girls		Boys			
	Mean (s.e.)	Std. Dev	Mean (s.e.)	Std. Dev	Mean (s.e.)	Std. Dev	t-value	d
Overall item format ability	.148172 (.0145241)	.7802559	.126920 (.0202117)	.7567919	.168248 (.0208067)	.8015302	-1.425	-0.053
Multiple-choice	.003556 (.0007088)	.0380775	-.000349 (.0010152)	.0380137	.007245 (.0009808)	.0377813	-5.381**	-0.200
Constructed-response	-.003563 (.0006950)	.0373341	.000238 (.0009948)	.0372483	-.007154 (.0009623)	.0370711	5.341**	0.199
Overall cognitive ability	.155506 (.0146812)	.7886988	.132959 (.0204126)	.7643143	.176808 (.0210459)	.8107442	-1.496	0.393
Knowing	.000110 (.0001747)	.0093857	-.000107 (.0002632)	.0098561	.000315 (.0002315)	.0089171	-1.206	-0.045
Applying	.000038 (.0001817)	.0097599	.000166 (.0002681)	.0100367	-.000083 (.0002464)	.0094928	.682	0.026
Reasoning	-.000148 (.0002943)	.0158101	-.000059 (.0004420)	.0165495	-.000232 (.0003915)	.0150834	.294	0.001

Note. ** indicates the value is significant at 0.01 level. * indicates the value is significant at 0.05 level.

Project Evaluation (計畫成果自評)

During the funding period (08/01/2015~ 07/31/2016), 1 journal paper and 1 international conference paper directly derived from this study have been on progress for submission and presented. Moreover, other 5 SSCI journal articles and 4 conference papers benefited from this MOST funding have been published, or presented. The publication list is presented as follows, and the contents of the underlined ones are derived from this study.

SSCI Journal Papers

0. **Liou, P.-Y.***, & Bulut, O. (submitted). Item format, cognitive domain, and gender interaction in TIMSS 2011 science results.
1. **Liou, P.-Y.***, & Ho, H. N. J. (accepted, July 2016). Relationships among instructional practices, students' motivational beliefs and science achievement in Taiwan using hierarchical linear modeling. *Research Papers in Education.* (SSCI)
2. Davenport, E. C., Davison, M. L., **Liou, P.-Y.**, & Love, Q. U. (2016). Easier said than done: Rejoinder on Sijtsma and on Green and Yang. *Educational Measurement: Issues and Practice, 35*(1), 6-10. (SSCI)
3. Davenport, E. C., Davison, M., **Liou, P.-Y.**, & Love, Q. (2015). Reliability, dimensionality, and internal consistency as defined by Cronbach: Distinct albeit related concepts. *Educational Measurement: Issues and Practice, 34*(4), 4-9. (SSCI)
4. **Liou, P.-Y.***, & Hung, Y.-C. (2015). Statistical techniques utilized in analyzing PISA and TIMSS databases in science education from 1996 to 2013: A methodological review. *International Journal of Science and Mathematics Education, 13*(6), 1449-1468. (SSCI)
5. **Liou, P.-Y.***, & Liu, E. Z, F. (2015). An analysis of the relationships between Taiwanese eighth and fourth graders' motivational beliefs and science achievement in TIMSS 2011. *Asia Pacific Education Review, 16*(3), 433-445. (SSCI)

*Corresponding author.

International Conferences

1. **Liou, P.-Y.**, & Wang, C.-L. (2016, Jun). *A content analysis of PISA and TIMSS studies from 1996 to 2015: The nexus of ILSAs and science education*. Paper presented at the 47th annual Australasian Science Education Research Association (ASERA) conference, Canberra, Australia.
2. **Liou, P.-Y.**, & Bulut, O. (2016, April). *Item format, cognitive domain, and gender interaction in TIMSS 2011 science results.* Paper presented at the 2016 annual meeting of American Educational Research Association, Washington, DC.
3. **Liou, P.-Y.** (2016, April). *Gender differences of motivational beliefs and science achievement in 26 countries.* Paper presented at the 2016 annual international conferences of National Association for Research in Science Teaching, Baltimore, MD.
4. Wang, C.-L., & **Liou, P.-Y.** (2016, April). *Taiwanese adolescents' motivational beliefs and*

science achievement: Evidence of TIMSS 2011. Paper presented at the 2016 annual international conferences of National Association for Research in Science Teaching, Baltimore, MD.

Local Conference

1. Wang, C.-L., & **Liou, P.-Y.** (2015, December). *Different patterns of motivational beliefs in science learning of the high and low performing students: Evidence of Taiwanese TIMSS 2011 data*. Paper presented at the 2015 annual international conference of Association of Science Education Taiwan (ASET), Kenting, Pingtung, Taiwan.

出國報告（出國類別：參加國際學術會議發表論文）

出席國際學術會議發表論文報告（研討會名稱：2016 Annual Meeting of American Educational Research Association）

服務機關：國立中央大學

姓名：劉佩艷

派赴國家：美國（華盛頓特區）

出國參與會議期間：105/04/08-105/04/12

報告日期：105/04/11

科技部補助國內專家學者出席國際學術會議報告

105 年 4 月 20 日

報告人姓名	劉佩艷	服務機構及職稱	國立中央大學 學習與教學研究所 助理教授
時間	4/8/2016~4/12/2016	地點	美國華盛頓特區
會議名稱	（中文）2016 年美國教育研究年會 （英文）2016 Annual Meeting of American Educational Research Association		
發表論文題目	（中文）TIMSS 2011 試題類型、認知階層與性別之交互作用 （英文）Item Format, Cognitive Domain, and Gender Interaction in TIMSS 2011 Science Results		

摘要

本報告的主要目的為描述報告人在美國華盛頓特區「美國教育研究協會」（American Education Research Association, 簡稱AERA）所舉辦的一年一度教育研究盛會的過程與相關紀錄。在此次會議中，除了發表了Item Format, Cognitive Domain, and Gender Interaction in TIMSS 2011 Science Results此研究，並與與會人士進行了學術交流外，也在許多場的演講與發表中獲取了許多對自己未來研究的相關重要知識與啟發。

目次

一、 目的

本文是參加2016 Annual Meeting of American Educational Research Association(AERA)之心得報告。除了參與自己的場次發表外，也參加了許多跟後學研究相關的場次以了解最新的研究趨勢，作為未來研究參考。

二、 參加會議經過和與會心得

此次的 2016 AERA 於華盛頓特區(Washington, D.C.)舉行，從 1916 年開始舉辦，此次為一百周年。報告人於 4/2 日搭乘大韓航空班機抵達坐落於 Virginia Arlington County 的 Ronald Reagan Washington National Airport (DCA)機場。此次會議，於 4 月 11 日由口頭報告形式發表一篇論文 Item Format, Cognitive Domain, and Gender Interaction in TIMSS 2011 Science Results。其餘時間積極地到有興趣的各場次聆聽演講，或是藉由海報發表與結構式海報發表與各研究者進行深度的互動討論。

此篇研究發表於 Division C – Learning and Instruction 中的 66.026 Assessing for Learning: Exploring Assessment Strategies for Supporting Conceptual Change。場次位於 Convection Center, Level One, Room 150 B，於 4:30-6:00pm 發表。因為此研究的量化技術較為高階，故跟於 University of Alberta 的 Okan Bulut 博士共同進行研究，Dr. Bulut 是測驗領域的優秀年輕學者，才剛在與 AERA 同時與舉辦的 National Council on Measurement in Education 國際會議中，因 Examining subscore reliability within Multidimensional IRT framework 得到了 Alicia Cascallar Award。被聆聽者被問的問題主要是此篇文章所使用的 the Rasch subdimension model 是如何用來分析因 booklet-design 所產生大量學生答案“遺失”等試題分析理論等技術性問題。此外，也有幾位聽眾評論說這篇研究呈現出統計上明顯的性別差異，即女生在建構式試題與高認知層次的題目表現的比男生好，等結果很值得發表。現場也有實際參與開發科學試題的研究者，他說雖然他只發展多選的題型，但若之後發展建構式試題，會注意可能產生的性別差異。

同場的還有“Using contextualized written tasks to assess differences in epistemological framing” by Brandy Buckingham from Northwestern University, “Secondary science teachers’ implementation of formative assessments in a learning progression-based environmental science curriculum” by Stacey Carpenter from

University of California-Santa Barbara, “Fostering model-based learning of human body systems during simulation-based investigations” by Barbara Buckley from WestEd, and “Effect of prior knowledge on inquiry in the 2009 NAEP science interactive computer task” by Jung Aa Moon from ETS.

除了發表自己的研究外，因為研究興趣，例如資料庫加值、科教政策與研究走向，去聆聽了兩場“研究與科學政策”(research and science policy)的特徵場次(feature sessions)，包含了 Future Directions for Longitudinal Studies Conducted by the National Center for Education Statistic 與 Research on Broadening Participation in STEM: Future Directions at the National Science Foundation。以及 “研究與科學政策”(Data use, access, and sharing)的兩場特徵場次，包含 The Contributions and Opportunities of Using Administration Data Systems in Research and Policy 與 Using Data Strategically and Effectively to Promote Learning Opportunities: A Dialogue Designed to Explore and Expand How We Think about Data use。從這幾場演講中，可以看得出美國教育學界與政府單位對於資料如何使用的更有加值效果著力尤深，此外，對於長期資料的蒐集與運用更是重點。

此外，我還參加了幾場由 SIG-Motivation in Education 所舉辦的場次，看到了好幾位知名的學習動機大師們，比如 Ann Renninger, Judith Harackiewicz, Allan Wigfield, Suzanne Hidi 等人。其中，對我最有直接收穫的是 “Mexican high school students’ motivational beliefs, self-regulatory behaviors, and academic achievement: A structural equation model” by David Chirinos 這一篇研究，作者使用了墨西哥 PISA 的資料來檢視這些變項之間的關係，其中一個發現是墨西哥學生的 utility value 比起 intrinsic value 對於學業成就更有解釋力，始發現跟文獻中的東亞學生類似，這是非常有趣的發現，因為西方學術主流的國家學生都是 intrinsic value 比起 utility value 來的更為重要。我自己用 TIMSS 的資料來看也是這樣，不過，在寫文章的時候，自己很難找到文獻可以引用，這個作者提供了下列的文獻，對之後的文獻引用極有幫助。此外，當作者認為文化是解釋這個現象時，特別可以由 collectivism 與 individualism 來解釋時，一位聽眾也回應說或許也可以往 Interpersonalism 與 Intrapersonalism 來解釋。

*The functioning of task interest value and utility value as predictors of effect have been found to differ between East Asian students and European-American students (Schecter, Durik, Miyamoto, & Haraackiewicz, 2011).

*Mexican culture, like East Asian cultures, generally described as collectivist, though East Asian students tend to be more motivated by utility value than interest value (Hong et al., 2009; Maddux & Yuki, 2006; Schecter et al., 2011).

在 “Why is writing about value so powerful?” by Hudith Harackiewicz 這篇文章中，她提到了 when students perceive value in academic tasks, they become more highly motivated, interested and engaged (Wigfield & Eccles, 1992; Eccles, Renninger & Hidi, 2016). They have been developing interventions to help students find value in their academic pursuits (Hulleman & Harackiewicz, 2009; Hulleman et al., 2010; Canning & Harackiewicz, 2015; Harackiewicz at al., 2016). These interventions focus on utility value as the most malleable of these task values. These interventions are most effective for at-risk students- those at risk of becoming disengaged, either because they have a history of poor performance or low interest.

在 “Fostering students’ value beliefs for mathematics with a relevance intervention in the classroom” by Hanna Gaspard 這篇文章中，她提到 students interest and value beliefs for mathematics decline throughout secondary school (e.g., Frenzel, Goetz, Pekrun, & Watt, 2010). Interventions based on expectancy-value theory has been applied to stimulate utility value and promote interest and engagement as an ultimate aim. Gaspard 在 2015 年於 Journal of Educational Psychology 針對 expectancy-value theory 發展了新的測量問卷以量測四大部分，包含 intrinsic value (e.g., I like doing math), attainment value (e.g., math is very important to me personally), utility value (e.g., I will often need math in my life), and cost (e.g., doing math makes me really nervous).

除了去各場次聽講外，我還非常喜歡去看海報發表，因為更有機會跟作者一對一對談，針對自己有興趣的部分直接詢問。在此紀錄幾場，在 “Level and pattern relationships: Beyond profile analysis” by Ernest Davenport 這篇研究中，作者強調 level mean 跟 group correlation 在分析資料的重要性，在了解到重要的 confounding variable (e.g., races or nations)是何者後，應該要加以控制，這樣的詮釋才是有意義且正確的。回到自己對 TIMSS 資料中跨國資料中學生學習信念的 paradoxical relationships 的研究中，這種方法或許是一種新方法來檢視與解釋這些關係。

Dr. Davenport 是我博士班的指導老師，過去在學業與生活上都受到他非常多的照顧。最近也有榮幸與他在 Educational Measurement: Issues and Practice 這個由 National Council on Measurement in Education 所支持的期刊上發表了兩篇探討 Chonbach’s alpha 的文章。

在 “The paradoxical relation of TIMSS gender differences in mathematics and the gender inequality index” by Thomas Hogan 這篇研究中，作者使用了由 the United Nations Human Development Programme 所編制的 the gender inequality index 來表示每個國家的性別不平等的狀況，這類的關於性別差異的跨國研究應有潛力，然而，由於台灣不是聯合國的成員之一，所以無法使用國際通用標準來指射台灣的狀況，甚為可惜。

三、 會議目錄

Item Format, Cognitive Domain, and Gender Interaction in TIMSS (Trends in International Mathematics and Science Study) 2011 Science Results. *Pey-Yan Liou, National Central University - Graduate Institute of Learning and Instruction; Okan Bulut, University of Alberta*

66.027. Learning, Instruction, and Motivation in Computer Science Education. Division C - Learning and Instruction; Paper Session
Convention Center, Level One, Room 145 A; 4:30-6:00pm

Chair: *Marie Bienkowski, SRI International*

Participants:

Designing Think-Aloud Interviews to Elicit Evidence of Computer Science Pedagogical Content Knowledge. *Alecia Hubbard, WestEd; Yvonne Kao, WestEd; Danielle Brown, WestEd*

Bringing Blocks-Based Programming Into High School Computer Science Classrooms. *David Weintrop, Northwestern University; Uri J. Wilensky, Northwestern University*

Coding Moves: Design-Based Research of Virtual Environment Interactions With Middle School Students. *Shaundra Bryant Daily, University of Florida; Alison E Leonard, Clemson University; Nikeetha Dsouza, Clemson University; Sabarish Babu, Clemson University; Sophie Joerg, Clemson University; Kara Gundersen, University of Florida; Dhaval Parmar, Clemson University; Xianshan Qu, Clemson University; Lorraine Lin*

Students' Initial Course Motivation and Their Achievement and Retention in Undergraduate Computer Science Courses. *Duane F. Shell, University of Nebraska - Lincoln; Abraham Flanigan, University of Nebraska - Lincoln; Markeya Peteranetz, University of Nebraska - Lincoln; Leen-Kiat Soh, University of Nebraska - Lincoln*

Assessing the Promise of the Exploring Computer Science Program. *Steven McGee, Northwestern University; Ronald I. Greenberg, Loyola University Chicago; Lucia Dettori, DePaul University; Dale F. Reed, University of Illinois at Chicago*

66.028. Promoting Public Understanding of Science: A Cornerstone of Diverse Democracies. Division C - Learning and Instruction; Symposium
Convention Center, Level Two, Room 207 A; 4:30-6:00pm

Chair: *Robert William Danielson, University of Southern California*

Participants:

Effects of Goal Priming on High School Students' Use of Mechanism and Evidence Information in a Science Media Text. *Jacqueline Wong, University of California - Los Angeles; William A. Sandoval, University of California - Los Angeles*

Cognitive Conflict Between Science and Intuition Across the Curriculum and Across the Life-Span. *Andrew Shtulman, Occidental College*

Four or So Ultrabrief Interventions That Increase Public Acceptance Regarding Global Warming. *Michael Ranney, University of California - Berkeley; Charles Chang, University of California - Berkeley; Tommy Ng, University of California - Berkeley; Justin Teicheira, University of California - Berkeley; Tina Luong, University of California - Berkeley; Lukas Gierth, University of California - Berkeley*

Overcoming Stumbling Blocks to Public Understanding of Science Through Refutation Texts and Graphics. *Gale M. Sinatra, University of Southern California; Robert William Danielson, University of Southern California*

Discussant: *Barbara K. Hofer, Middlebury College*

66.029. Relations Between Motivation and Anxiety Beliefs on Mathematical Achievement. Division C - Learning and Instruction; Paper Session
Convention Center, Level One, Room 144 B; 4:30-6:00pm

Chair: *Sheila R. Vaidya, Drexel University*

Participants:

The Intergenerational Transmission of Math Anxiety. *Wondimu Ahmed, The University of Akron*

Identify Profiles of At-Risk Middle-Grade Students in Mathematics in an Urban School. *Mei-Lin Chang, Kennesaw State University; Melanie M. Keller, University of Konstanz; Shanna Ricketts, Emory University; Morgan Z-J Faison, Emory University; Patricia Vela, Emory University*

The Development of Situational Interest in Math: Effects of Triggered Situational Interest and Utility Value. *Sungjun Won, The Ohio State University; Shirley L. Yu, The Ohio State University*

What Keeps Chinese Students Motivated in Doing Math Homework? An Empirical Investigation. *Ruiping Yuan, Mississippi State University; Jianzhong Xu, Mississippi State University; Fuyi Yang, East China Normal University; Judy Randi, University of New Haven; Weiya Wang, Qilu Normal University; Wei Wang, Qilu Normal University; Yan Sun, Qilu Normal University*

Finding the Answers Behind Students' Confidence Level in Mathematics Across Countries: A Secondary Analysis Using TIMSS (Trends in International Mathematics and Science Study) 2011. *Linh Nguyet Doan, Teachers College, Columbia University*

66.030. Social and Cultural Influences on Motivation and Engagement. Division C - Learning and Instruction; Paper Session
Convention Center, Level Two, Room 209 A; 4:30-6:30pm

Chair: *Jessica J. Summers, The University of Arizona*

Participants:

Social Context as a Motivational Resource in School: Implications for Academic Adjustment. *Gwen C. Marchand, University of Nevada - Las Vegas; Carrie Furrer, Portland Public Schools*

Interdependence of Perceived Peer Motivational Climate and Social Goal Orientations Among Middle and High School Students. *Efrat Mezan, Bar-Ilan University; Shir Ester Rubin, Bar-Ilan University; Melissa Karakus, Temple University; Nir Madjar, Bar Ilan University*

The Effects of Parental Control on Adolescents' Motivational Outcomes. *EunJin Seo, The University of Texas - Austin; Erika Alisha Patall, The University of Texas - Austin; Marilla D. Svinicki, The University of Texas - Austin*

Testing the Relevance of Expectancy-Value Theory in the Caribbean Context. *Tamica G. Martin, University of New South Wales; Paul Evans, School of Education; Andrew J. Martin, The University of New South Wales*

A Chinese Perspective on Self-Regulated Learning and Motivation: An Interview Study. *Jing Wang, The University of Hong Kong; Jingyan Lu, The University of Hong Kong*

66.031. The Role of Teacher Practice in Promoting Academically Productive Student Dialogue: Past, Present, and Future. Division C - Learning and Instruction; Symposium
Convention Center, Level One, Room 150 A; 4:30-6:00pm

Chair: *Noreen M. Webb, University of California - Los Angeles*

Participants:

Promoting Academically Productive Student Dialogue. *Annemarie S. Palincsar, University of Michigan; Meredith Baker, University of Michigan - Ann Arbor; Miranda Fitzgerald, University of Michigan - Ann Arbor; Carrie-Anne Sherwood, University of Michigan*

Supporting Teachers in Taking Up Productive Talk Moves: Challenges of Professional Learning at Scale. *Catherine O'Connor, Boston University; Sarah Michaels, Clark University*

Teacher Practices That Promote Productive Dialogue and Learning in Mathematics Classrooms. *Noreen M. Webb, University of California - Los Angeles; Megan L. Franke, University of California - Los Angeles; Marsha M. Ing, University of California - Riverside; Angela Chan Turrou, University of California - Los Angeles; Nicholas Charles Johnson, University of California - Los Angeles; Joy Zimmerman, University of California - Los Angeles*

Promoting Academically Productive Student Dialogue: A Personal Journey. *Robyn Margaret Gillies, The University of Queensland*

Discussant: *Frederick D. Erickson, University of California - Los Angeles*

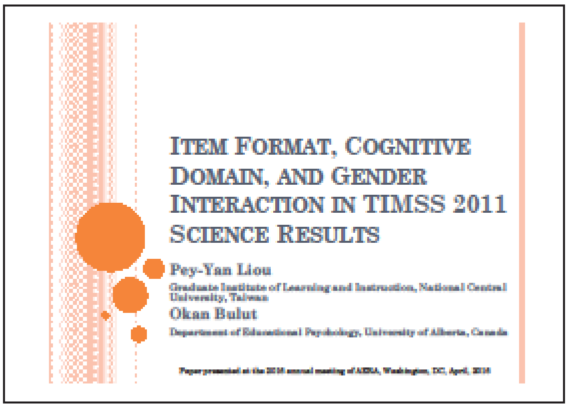
66.032. Toward Building Makerspaces for All: New Theories and Practices to Design Inclusive Makerspaces. Division C - Learning and Instruction; Structured Poster Session
Convention Center, Level One, Room 101; 4:30-6:00pm

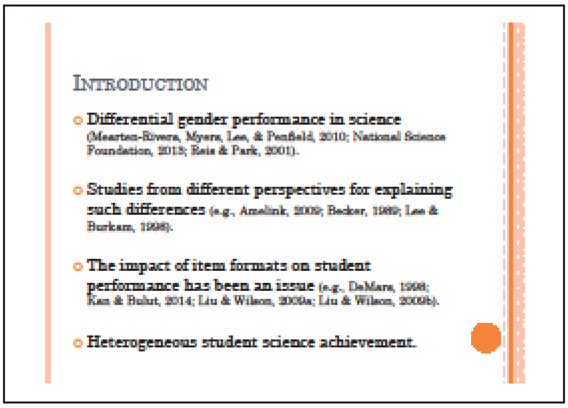
Chair: *Angela Calabrese Barton, Michigan State University*

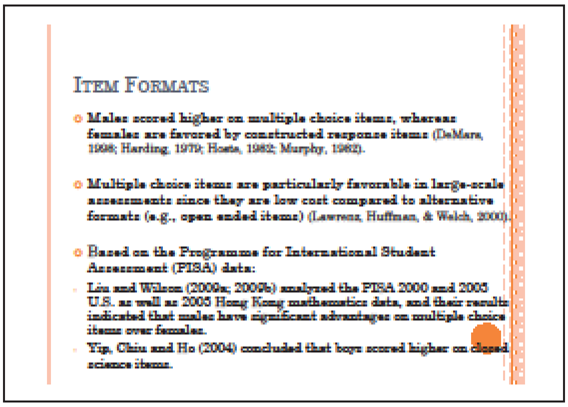
Participants:

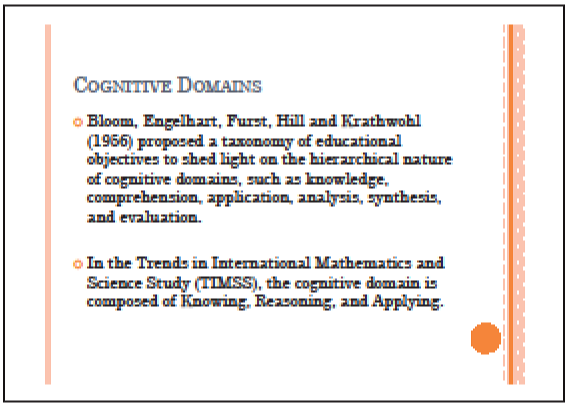
1. Resourceful and Inclusive: Toward Design Principles for Makerspaces. *Kimberly Marie Sheridan, George Mason University; Abigail W Konopasky, George Mason University; Asia Williams, Boys and Girls Clubs of Greater Washington; Grace Jessica Juanita Wingo, George Mason University*
2. Exploring How a "Judgment-Free" Makerspace Supports African American Girls' Identity Work as Community Makers and Engineers. *Faith Brown Freeman, University of North Carolina - Greensboro; Edna Tan, University of North Carolina at Greensboro*
3. Youth Engagement and Mobilities of Learning During Making in an Equity-Oriented Makerspace. *Myunghwan Shin, Michigan State*

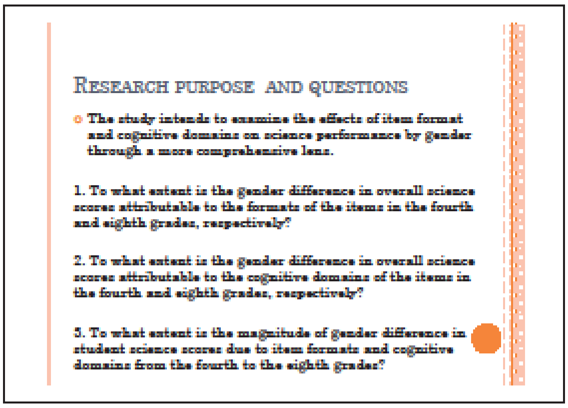
四、　附錄一：報告簡報

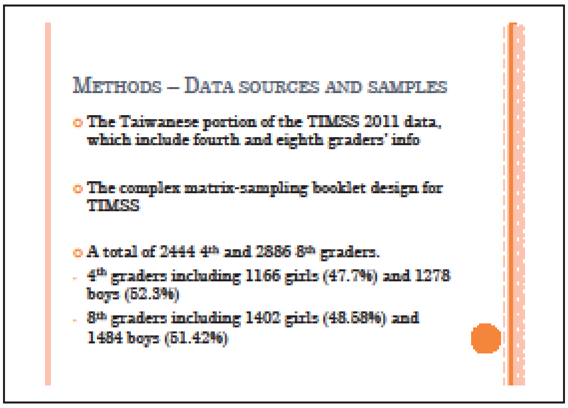

ITEM FORMAT, COGNITIVE DOMAIN, AND GENDER INTERACTION IN TIMSS 2011 SCIENCE RESULTS
Pey-Yan Liou
Graduate Institute of Learning and Instruction, National Central University, Taiwan
Okan Bulut
Department of Educational Psychology, University of Alberta, Canada


INTRODUCTION
Differential gender performance in science (Mearten-Rivera, Myers, Lee, & Penfield, 2010; National Science Foundation, 2013; Reis & Park, 2001).
Studies from different perspectives for explaining such differences (e.g., Amelink, 2009; Becker, 1989; Lee & Burkam, 1996).
The impact of item formats on student performance has been an issue (e.g., DeMars, 1998; Kan & Bulut, 2014; Liu & Wilson, 2009a; Liu & Wilson, 2009b).
Heterogeneous student science achievement.


ITEM FORMATS
Males scored higher on multiple choice items, whereas females are favored by constructed response items (DeMars, 1998; Harding, 1979; Hoste, 1982; Murphy, 1982).
Multiple choice items are particularly favorable in large-scale assessments since they are low cost compared to alternative formats (e.g., open ended items) (Lawrenz, Huffman, & Welch, 2000).
Based on the Programme for International Student Assessment (PISA) data:
Liu and Wilson (2009a; 2009b) analyzed the PISA 2000 and 2003 U.S. as well as 2003 Hong Kong mathematics data, and their results indicated that males have significant advantages on multiple choice items over females.
Yip, Chiu and Ho (2004) concluded that boys scored higher on closed science items.


COGNITIVE DOMAINS
Bloom, Engelhart, Furst, Hill and Krathwohl (1956) proposed a taxonomy of educational objectives to shed light on the hierarchical nature of cognitive domains, such as knowledge, comprehension, application, analysis, synthesis, and evaluation.
In the Trends in International Mathematics and Science Study (TIMSS), the cognitive domain is composed of Knowing, Reasoning, and Applying.


RESEARCH PURPOSE AND QUESTIONS
The study intends to examine the effects of item format and cognitive domains on science performance by gender through a more comprehensive lens.
1. To what extent is the gender difference in overall science scores attributable to the formats of the items in the fourth and eighth grades, respectively?
2. To what extent is the gender difference in overall science scores attributable to the cognitive domains of the items in the fourth and eighth grades, respectively?
3. To what extent is the magnitude of gender difference in student science scores due to item formats and cognitive domains from the fourth to the eighth grades?


METHODS – DATA SOURCES AND SAMPLES
The Taiwanese portion of the TIMSS 2011 data, which include fourth and eighth graders' info
The complex matrix-sampling booklet design for TIMSS
A total of 2444 4th and 2886 8th graders.
4th graders including 1166 girls (47.7%) and 1278 boys (52.3%)
8th graders including 1402 girls (48.58%) and 1484 boys (51.42%)

Methods – Measures

- Four variables: 1) item formats, 2) cognitive domains, 3) students' science achievement scores by specific categories, and 4) gender.
- 72 out of 172 items for the 4th graders and 88 out of 217 items for the 8th graders

Distribution of Item Format and Cognitive Domains

4th grade items	Knowing	Applying	Reasoning	Total
Multiple choice	20	18	6	44
Constructed response	10	13	5	28
Total	30	31	11	72

8th grade items	Knowing	Applying	Reasoning	Total
Multiple choice	22	15	7	44
Constructed response	10	19	15	44
Total	32	34	22	88

Methods – Statistical analyses

- The Rasch subdimension model (Brandt, 2008).

 - The model is a special case of a bi-factor Rasch model by incorporating an additional set of parameters for estimating an overall ability and subscores for each subdimension.

- Independent-sample t tests

Results – 4th graders

	Total		Girls		Boys		
	Mean (s.e.)	Std. Dev	Mean (s.e.)	Std. Dev	Mean (s.e.)	Std. Dev	t-value
Overall item format ability	.345626 (.01058)	.523	.291272 (.01489)	.506	.398217 (.01487)	.531	-4.936**
Multiple choice	.073085 (.00223)	.110	.063183 (.00314)	.107	.082023 (.00313)	.112	-4.242**
Constructed response	-.157274 (.00480)	.237	-.136130 (.00676)	.231	-.178564 (.00675)	.241	4.229**
Overall cognitive ability	.422154 (.01292)	.638	.357285 (.01820)	.621	.481336 (.1815)	.648	-4.825**
Knowing	.000860 (.00014)	.007	.000258 (.00020)	.007	.001409 (.00019)	.007	-4.055**
Applying	.000285 (.00013)	.006	.000548 (.00018)	.006	.000045 (.00019)	.006	1.865
Reasoning	-.001145 (.00015)	.007	-.000806 (.00022)	.007	-.001455 (.00021)	.007	2.096*

Results – 8th graders

	Total		Girls		Boys		
	Mean (s.e.)	Std. Dev	Mean (s.e.)	Std. Dev	Mean (s.e.)	Std. Dev	t-value
Overall item format ability	.148172 (.01452)	.780	.126920 (.02021)	.756	.168248 (.02080)	.801	-1.425
Multiple choice	.003656 (.00070)	.038	-.000349 (.00101)	.038	.007245 (.00098)	.037	-5.361**
Constructed response	-.003563 (.00069)	.037	.000238 (.00099)	.037	-.007154 (.00096)	.037	5.341**
Overall cognitive ability	.155506 (.01468)	.788	.132959 (.02041)	.764	.176806 (.02104)	.810	-1.496
Knowing	.000110 (.00017)	.009	-.000107 (.00026)	.009	.000315 (.00023)	.008	-1.206
Applying	.000036 (.00018)	.009	.000168 (.00026)	.010	-.000083 (.00024)	.009	.682
Reasoning	-.000148 (.00029)	.015	-.000059 (.00044)	.016	-.000232 (.00039)	.015	.294

Conclusions

- Item format:
  - Girls had a significant disadvantage compared to boys in multiple choice items, but had a significant advantage in constructed response items in both grades while the overall item format on gender differences may not differ.
  - The gender gap seems to be wider in the 8th grade than in the 4th grade.
- Cognitive domain:
  - Boys performed better in lower-level cognitive domains than girls.
  - The gap of gender differences seems to be narrower in the 8th grade than in the 4th grade.
- In sum: Due to the more percentages of items on constructed response and cognitive domains on the 8th grade exam, no significant gender difference is found.

Implications

- The study shed light on the patterns and particular gender differences to identify the two factors that are likely to introduce statistical differences.
- The results could serve as guidelines for the development of achievement testing when designing gender-based initiatives to address gender differences in science achievement.

Limitations

- Only students' responses to science items, but not other contextual variables, were used to estimate students' science scores in each domain of item formats and cognitive levels.
- The dichotomous categorization of item formats (i.e., multiple-choice and constructed response items) may be too simplified to well represent the formats of items.

Thank you for listening.

Any comments or questions?

Pey-Yan Liou: lioupeyyan@gmail.com
Okan Bulut: bulut@ualberta.ca

Acknowledgements: the Ministry of Science and Technology 104-[illegible] & the [illegible] funding from the China Foundation for the Promotion of Education and Culture

科技部補助計畫衍生研發成果推廣資料表

日期:2016/10/30

科技部補助計畫	計畫名稱: 以TIMSS 2011台灣學生科學成就資料檢視試題類型、內容領域與認知階層中的性別差異(A06)
	計畫主持人: 劉佩艷
	計畫編號: 104-2629-S-008-001- 學門領域: 性別與科技研究

無研發成果推廣資料

104年度專題研究計畫成果彙整表

計畫主持人：劉佩艶	計畫編號：104-2629-S-008-001-

計畫名稱：以TIMSS 2011台灣學生科學成就資料檢視試題類型、內容領域與認知階層中的性別差異(A06)

成果項目					量化	單位	質化（說明：各成果項目請附佐證資料或細項說明，如期刊名稱、年份、卷期、起訖頁數、證號...等）
國內	學術性論文	期刊論文			0	篇	
國內	學術性論文	研討會論文			1	篇	Wang, C.-L., & Liou, P.-Y. (2015, December). Different patterns of motivational beliefs in science learning of the high and low performing students: Evidence of Taiwanese TIMSS 2011 data. Paper presented at the 2015 annual international conference of Association of Science Education Taiwan (ASET), Kenting, Pingtung, Taiwan.
國內	學術性論文	專書			0	本	
國內	學術性論文	專書論文			0	章	
國內	學術性論文	技術報告			0	篇	
國內	學術性論文	其他			0	篇	
國內	智慧財產權及成果	專利權	發明專利	申請中	0	件	
國內	智慧財產權及成果	專利權	發明專利	已獲得	0	件	
國內	智慧財產權及成果	專利權	新型/設計專利		0	件	
國內	智慧財產權及成果	商標權			0	件	
國內	智慧財產權及成果	營業秘密			0	件	
國內	智慧財產權及成果	積體電路電路布局權			0	件	
國內	智慧財產權及成果	著作權			0	件	
國內	智慧財產權及成果	品種權			0	件	
國內	智慧財產權及成果	其他			0	件	
國內	技術移轉	件數			0	件	
國內	技術移轉	收入			0	千元	
國外	學術性論文	期刊論文			5	篇	1. Relationships among instructional practices, students' motivational beliefs and science achievement in Taiwan using hierarchical linear modeling. Research Papers in Education. 2. Easier said than done: Rejoinder on Sijtsma and on Green and Yang. Educational Measurement: Issues and

							Practice, 35(1), 6-10. 3. Reliability, dimensionality, and internal consistency as defined by Cronbach: Distinct albeit related concepts. Educational Measurement: Issues and Practice, 34(4), 4-9. 4. Statistical techniques utilized in analyzing PISA and TIMSS databases in science education from 1996 to 2013: A methodological review. International Journal of Science and Mathematics Education, 13(6), 1449-1468. 5. An analysis of the relationships between Taiwanese eighth and fourth graders' motivational beliefs and science achievement in TIMSS 2011. Asia Pacific Education Review, 16(3), 433-445.
		研討會論文			4		
		專書			0	本	
		專書論文			0	章	
		技術報告			0	篇	
		其他			0	篇	
	智慧財產權及成果	專利權	發明專利	申請中	0	件	
				已獲得	0		
			新型/設計專利		0		
		商標權			0		
		營業秘密			0		
		積體電路電路布局權			0		
		著作權			0		
		品種權			0		
		其他			0		
	技術移轉	件數			0	件	
		收入			0	千元	
參與計畫人力	本國籍	大專生			1	人次	
		碩士生			3		
		博士生			0		
		博士後研究員			0		
		專任助理			0		
	非本國籍	大專生			0		
		碩士生			0		
		博士生			0		

		博士後研究員	0		
		專任助理	0		
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。）					

	成果項目	量化	名稱或內容性質簡述
科教國合司計畫加填項目	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	

科技部補助專題研究計畫成果自評表

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（簡要敘述成果是否具有政策應用參考價值及具影響公共利益之重大發現）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（請於其他欄註明專利及技轉之證號、合約、申請及洽談等詳細資訊） 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200字為限）
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） The purpose of this study was to examine 4th and 8th grade students' science performance in terms of the interaction between gender and two test-related components, item format and cognitive domains. The results of this study shed light the possible differential function of the item format on student science achievement by gender.
4. 主要發現 本研究具有政策應用參考價值：□否 ■是，建議提供機關教育部 （勾選「是」者，請列舉建議可提供施政參考之業務主管機關） 本研究具影響公共利益之重大發現：■否 □是 說明：（以150字為限） The impact of item format on student science achievement enlarges from the 4th to the 8th grade.