การวิเคราะห์เปรียบเทียบโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์: การประยุกต์ใช้โมเดลมูลค่าเพิ่มพหุระดับที่มีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ

นายพนัส จันทร์เปล่ง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)
เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ ที่ส่งผ่านทางบัณฑิตวิทยาลัย
The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository (CUIR)
are the thesis authors' files submitted through the University Graduate School.

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาครุศาสตรดุษฎีบัณฑิต
สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา
คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2557
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

A COMPARATIVE ANALYSIS OF QUALITY ASSESSMENT MODELS OF SCIENCE INSTRUCTIONAL MANAGEMENT: AN APPLICATION OF MULTILEVEL VALUE-ADDED MODEL WITH DIFFERENTIAL ITEM FUNCTIONING AND TEST FUNCTIONING ANALYSES

Mr. Panat Chanpleng

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in Educational Measurement and Evaluation
Department of Educational Research and Psychology
Faculty of Education
Chulalongkorn University
Academic Year 2014

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การวิเคราะห์เปรียบเทียบโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์: การประยุกต์ใช้โมเดลมูลค่าเพิ่มพหุระดับที่มีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ
โดย	นายพนัส จันทร์เปล่ง
สาขาวิชา	การวัดและประเมินผลการศึกษา
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ผู้ช่วยศาสตราจารย์ ดร. ณัฏฐภรณ์ หลาวทอง
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม	ผู้ช่วยศาสตราจารย์ ดร. สังวรณ์ งัดกระโทก

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาดุษฎีบัณฑิต

.....คณบดีคณะครุศาสตร์
(รองศาสตราจารย์ ดร. บัญชา ชลาภิรมย์)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ
(ศาสตราจารย์ ดร. ศิริชัย กาญจนวาสี)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(ผู้ช่วยศาสตราจารย์ ดร. ณัฏฐภรณ์ หลาวทอง)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม
(ผู้ช่วยศาสตราจารย์ ดร. สังวรณ์ งัดกระโทก)

.....กรรมการ
(รองศาสตราจารย์ ดร. โชติกา ภาษีผล)

.....กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร. กมลวรรณ ตังธนกานนท์)

.....กรรมการภายนอกมหาวิทยาลัย
(ดร. ชูศักดิ์ ขัมภลิขิต)

พนัส จันทร์เปล่ง : การวิเคราะห์เปรียบเทียบโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์: การประยุกต์ใช้โมเดลมูลค่าเพิ่มพหุระดับที่มีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ (A COMPARATIVE ANALYSIS OF QUALITY ASSESSMENT MODELS OF SCIENCE INSTRUCTIONAL MANAGEMENT: AN APPLICATION OF MULTILEVEL VALUE-ADDED MODEL WITH DIFFERENTIAL ITEM FUNCTIONING AND TEST FUNCTIONING ANALYSES) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: ผศ. ดร. ณัฏฐภรณ์ หลาวทอง, อ.ที่ปรึกษาวิทยานิพนธ์ร่วม: ผศ. ดร. สังวรณ์ งัดกระโทก, 276 หน้า.

การวิจัยครั้งนี้มีจุดมุ่งหมายเพื่อ 1) เพื่อศึกษาผลการทำหน้าที่ต่างกันของข้อสอบ (DIF) ที่มีการตรวจให้คะแนนทวิวิภาคและพหุวิภาค และการทำหน้าที่ต่างกันของแบบทดสอบ (DTF) ใน แบบทดสอบการประเมินการรู้เรื่องวิทยาศาสตร์ 2) เพื่อเปรียบเทียบประสิทธิภาพของโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์เมื่อใช้โมเดลมูลค่าเพิ่มพหุระดับ 2 โมเดล คือ โมเดลที่ไม่ได้พิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ (โมเดลที่ 1) กับโมเดลที่มีการพิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบโดยการตัดข้อสอบที่มีการทำหน้าที่ต่างกันของข้อสอบและการทำหน้าที่ต่างกันเป็นขั้นที่แตกต่างกัน (โมเดลที่ 2) ซึ่งในการเปรียบเทียบประสิทธิภาพของโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ มีวัตถุประสงค์ย่อยสองข้อ คือ 2.1) เพื่อศึกษาความสัมพันธ์ระหว่างคะแนนมูลค่าเพิ่มกับ 3 ตัวแปร คือ เศรษฐานะของครอบครัว แหล่งทรัพยากรการศึกษาที่บ้านและความมั่งคั่งของครอบครัว ทั้งก่อนและหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน และ 2.2) เพื่อเปรียบเทียบความสอดคล้องของการจัดอันดับคุณภาพและการจัดกลุ่มคุณภาพของโมเดลมูลค่าเพิ่มพหุระดับทั้ง 2 โมเดล ข้อมูลที่ใช้ในการวิจัยครั้งนี้เป็นข้อมูลทุติยภูมิจากโครงการประเมินผลนักเรียนนานาชาติเกี่ยวกับการรู้เรื่องวิทยาศาสตร์ (PISA) ปี ค.ศ. 2009 จำนวน 9 ฉบับ ของกลุ่มตัวอย่างในประเทศไทย ซึ่งประกอบด้วยนักเรียนจำนวน 4,292 คน จากสถานศึกษา 230 แห่ง ผลการวิจัยที่สำคัญพบว่า

1) การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ พบว่า มีข้อสอบที่เอนเอียงเข้าข้างนักเรียนหญิง (3 ข้อ), นักเรียนที่เรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา (1 ข้อ), นักเรียนที่มาจากครอบครัวที่มีเศรษฐานะต่ำ (2 ข้อ), นักเรียนที่มาจากครอบครัวที่มีแหล่งทรัพยากรที่บ้านสูง (5 ข้อ), และนักเรียนที่มาจากครอบครัวที่มีความมั่งคั่งสูง (2 ข้อ) ค่าความเที่ยงของแบบทดสอบในโมเดลที่ 2 ลดลงเมื่อข้อสอบที่ตรวจพบว่าทำหน้าที่ต่างกันถูกตัดออก ส่วนผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบ พบว่า ขนาดอิทธิพลการทำหน้าที่ต่างกันของแบบสอบ 2 ฉบับ ลดลงหลังจากตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันออกไป ส่วนแบบสอบอีก 7 ฉบับมีขนาดอิทธิพลการทำหน้าที่ต่างกันของแบบสอบไม่เปลี่ยนแปลง และหลังจากตัดข้อสอบบางข้อที่ตรวจพบการทำหน้าที่ต่างกันออกไปในโมเดลที่ 2 พบว่า ผลการวิเคราะห์องค์ประกอบเชิงยืนยันทั้งแบบสอบรวมและแยกรายฉบับมีความตรงเชิงโครงสร้างที่ยอมรับได้

2) ผลการศึกษาสัมประสิทธิ์การทำนาย (R^2) พบว่า โมเดลที่ 1 มีประสิทธิภาพการทำนายมากกว่าโมเดลที่ 2 โดยที่สัมประสิทธิ์การทำนาย (R^2) ของโมเดลที่ 1 เท่ากับ .49946 ซึ่งสูงกว่า โมเดลที่ 1 (R^2 เท่ากับ .49507)

2.1) ผลการศึกษาความสัมพันธ์ระหว่างคะแนนมูลค่าเพิ่มกับ 3 ตัวแปรควบคุม พบว่า คะแนนมูลค่าเพิ่มกับทั้ง 3 ตัวแปรมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และคะแนนมูลค่าเพิ่มกับทั้ง 3 ตัวแปรภายหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันออกไปแล้วในการวิเคราะห์มูลค่าเพิ่มมีขนาดลดลงเมื่อเปรียบเทียบกับการวิเคราะห์ที่ไม่ได้ตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันออกไป

2.2) ผลการเปรียบเทียบความสอดคล้องของการจัดอันดับคุณภาพโรงเรียนจากคะแนนมูลค่าเพิ่มโดยสถิติทดสอบ Wilcoxon Signed Ranks พบว่า โมเดลที่ 1 และโมเดลที่ 2 มีการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่มไม่สอดคล้องกัน ส่วนผลการเปรียบเทียบความสอดคล้องของการจัดกลุ่มคุณภาพโรงเรียน 5 กลุ่ม ตามคะแนนมูลค่าเพิ่ม พบว่า โมเดลที่ 1 กับโมเดลที่ 2 มีการจัดกลุ่มคุณภาพโรงเรียนที่สอดคล้องกันอย่างมีนัยสำคัญทางสถิติ (Cohen's Kappa = .865, $p<.05$)

ภาควิชา	วิจัยและจิตวิทยาการศึกษา	ลายมือชื่อนิสิต .....
สาขาวิชา	การวัดและประเมินผลการศึกษา	ลายมือชื่อ อ.ที่ปรึกษาหลัก .....
ปีการศึกษา	2557	ลายมือชื่อ อ.ที่ปรึกษาร่วม .....

5384236327 : MAJOR EDUCATIONAL MEASUREMENT AND EVALUATION
KEYWORDS: DIFFERENTIAL ITEM FUNCTIONING (DIF) / DIFFERENTIAL TEST FUNCTIONING (DTF) / DIFFERENTIAL STEP FUNCTIONING (DSF) / VALUE-ADDED ANALYSIS

PANAT CHANPLENG: A COMPARATIVE ANALYSIS OF QUALITY ASSESSMENT MODELS OF SCIENCE INSTRUCTIONAL MANAGEMENT: AN APPLICATION OF MULTILEVEL VALUE-ADDED MODEL WITH DIFFERENTIAL ITEM FUNCTIONING AND TEST FUNCTIONING ANALYSES. ADVISOR: ASST. PROF. NUTTAPORN LAWTHONG, Ph.D., CO-ADVISOR: ASST. PROF. SUNGWON NGUDGRATOKE, Ph.D., 276 pp.

The purposes of this research were to 1) examine the differential item functioning (DIF) and differential test functioning (DTF) of dichotomous and polytomously scored science literacy items, 2) compare the efficiency of the two value-added models between the model without the inclusion of differential item functioning and differential test functioning (Model1) and the model with the inclusion of differential item functioning and differential test functioning (Model 2). For the Model 2, DIF and DTF analyses were used to identify flawed items and subsequently items identified as DIF or differential step functioning (DSF) were deleted and not used in the value-added model. The efficiency of the models was evaluated by the correlation between value-added values provided by Model1 and Model2 and the three control variables: socioeconomic status, home educational resources, and family wealth, the consistency between rankings of value-added values from the two models and the consistency of school classifications based on the value-added values from the two models. Data used in this study were the secondary data of Thailand's 2009 student literacy assessment which were from Programme for International Students Assessment (PISA). Data were consisted of 9 booklets and responses of 4,292 students in 230 schools throughout Thailand were used. The major research findings were found as follows:

1) The differential item functioning analysis indicated that there were items identified as DIF items favoring females (3 items), tutoring students beyond science classrooms (1 item), low-socioeconomic status students (2 items), students with high home educational resources (5 items), and students from high wealth family (2 items). The reliability coefficient from Model 2 decreased as the DIF items were deleted. Differential test functioning (DTF) analysis of the data for the two booklets showed that the magnitude of DTF decreased as DIF items were deleted, while the magnitudes of DTF were unchanged for other seven booklets. After some DIF items in the Model 2 were removed, confirmatory factor analysis was applied to the total test and to individual booklet, and it was found that construct validity evidences were acceptable.

2) Model 1 was more efficient than Model 2 in terms of proportion of variance explained (R^2); R^2 for Model 1 was .49946 which was higher than R^2 for Model2 (.49507).

2.1) The correlation between value-added measures and three control variables were statistically significant at .05 and getting smaller as DIF items were deleted in the value-added estimation process compared with when DIF items were not deleted.

2.2) The Wilcoxon Signed Ranks Test applied to two school ranking data based on the ranking of school value-added measures from the Model 1 and the Model 2 was not statistically significant. However; when schools were classified into five grades based on their value-added measures from Model 1 and Model 2, there was a statistically consistent between school classifications provided by the two model (Cohen's Kappa = .865, p <.05).

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Department:	Educational Research and Psychology	Student's Signature	.....
Field of Study:	Educational Measurement and Evaluation	Advisor's Signature	.....
Academic Year:	2014	Co-Advisor's Signature	.....

กิตติกรรมประกาศ

วิทยานิพนธ์เล่มสำเร็จลุล่วงเป็นไปด้วยดี ด้วยความเมตตากรุณาอย่างยิ่งจาก ผู้ช่วยศาสตราจารย์ ดร.ณัฏฐภรณ์ หลาวทอง อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก และผู้ช่วยศาสตราจารย์ ดร.สังวรณ์ งัดกระโทก อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม ที่ได้คำแนะนำที่เป็นประโยชน์และข้อเสนอแนะในการปรับปรุงแก้ไขวิทยานิพนธ์เล่มนี้ และขอขอบพระคุณ Professor Dr. Randy Penfield ผู้จุดประกายในการศึกษาวิชาการวัดผลในด้านการทำหน้าที่ต่างกันเป็นขั้นและการทำหน้าที่ต่างกันของแบบสอบ อีกทั้งเป็นผู้ที่ให้ความอนุเคราะห์ให้ใช้โปรแกรมวิเคราะห์ข้อมูล DIFAS

กราบขอบพระคุณ ศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี รองศาสตราจารย์ ดร.ศิริเดช สุขชีวะ รองศาสตราจารย์ ดร.โชติกา ภาษีผล และผู้ช่วยศาสตราจารย์ ดร.กมลวรรณ ตังธนกานนท์ ที่ให้ข้อเสนอแนะและคำปรึกษาที่เป็นประโยชน์ในวิชาสัมมนาสาขาการวัดและประเมินผลการศึกษา และกราบขอบพระคุณคณะกรรมการสอบวิทยานิพนธ์ทุกท่านรวมถึงกรรมการภายนอก ได้แก่ อาจารย์ ดร.ชูศักดิ์ ขัมภลิขิต ที่ให้คำแนะนำที่มีคุณค่าและมีประโยชน์ต่อการปรับปรุงวิทยานิพนธ์ให้สมบูรณ์ อีกทั้งคณาจารย์ภาควิชาวิจัยและจิตวิทยาทุกท่านที่ประสิทธิ์ประสาทวิชาความรู้ ความสามารถที่เป็นอยู่ทุกวันนี้ ซึ่งข้าพเจ้ารู้สึกประทับใจและซาบซึ้งในพระคุณอย่างสูงของทุกๆ ท่าน เรื่อยมาจนถึงทุกวันนี้

ขอขอบพระคุณสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (สสวท.) ที่ให้ข้อมูลโครงการประเมินผลนักเรียนนานาชาติ (PISA) และบัณฑิตวิทยาลัยที่ให้การสนับสนุน ทุน 90 ปี จุฬาลงกรณ์มหาวิทยาลัยเพื่อนำมาใช้ในการศึกษาวิจัยครั้งนี้ อีกทั้ง ดร.ประกฤติยา ทักษิโณ และดร.เรืองเดช ศิริกิจ ที่ให้คำปรึกษาในวิทยานิพนธ์ และเพื่อนร่วมสาขาการวัดและประเมินผลทุกๆ ท่าน เพื่อนที่สนิท และมิตรสหายทั้งทางวิชาการและไม่ใช่วิชาการที่ให้คำแนะนำในการปรับปรุงวิทยานิพนธ์ให้สำเร็จลุล่วงเป็นไปด้วยดี

กราบขอบพระคุณพระราชพรหมยาน หลวงพ่อเกษม พระครูวิลาศกาญจนธรรม พระครูปลัดสมปอง และหลวงอาฉัตรเวช ผู้ซึ่งเป็นเสมือนครูคนแรกที่คอยเตือนสติของข้าพเจ้าในระหว่างใช้ชีวิตทั้งในระหว่างเรียนและทำงานให้สำเร็จลุล่วงมาจนถึงทุกวันนี้ นอกจากนี้ข้าพเจ้าขอกราบขอบพระคุณคุณพ่อพนมศักดิ์และคุณแม่ทองล้วน จันทร์เปล่ง และคุณครูทิพย์วัลย์ จันทร์เปล่ง ที่เป็นกำลังสำคัญและเป็นแรงผลักดันให้ข้าพเจ้าสำเร็จการศึกษาในระดับปริญญาเอก ตลอดจนญาติๆ ทุกท่านที่คอยสนับสนุนและให้กำลังใจแก่ข้าพเจ้าตลอดมา ท้ายที่สุดนี้ ข้าพเจ้าหวังว่างานวิจัยดังกล่าวจะเป็นประโยชน์และเพิ่มพูนความรู้ให้ท่านผู้ที่สนใจ เพื่อใช้เป็นแนวทางในการพัฒนางานวิจัยอันจะพึงสร้างคุณประโยชน์ต่อวงการศึกษาในทุกๆ ระดับ และทุกๆ ด้าน เพื่อความเจริญงอกงามทางด้านการศึกษาที่จะส่งผลไปยังประเทศชาติต่อไปในอนาคต

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ฉ

สารบัญ ..... ช

สารบัญตาราง ..... ฐ

สารบัญรูปภาพ ..... ด

บทที่ 1 บทนำ 1

ความเป็นมาและความสำคัญของปัญหา 1

คำถามการวิจัย 11

วัตถุประสงค์การวิจัย 11

สมมติฐานการวิจัย 12

ขอบเขตการวิจัย 13

คำจำกัดความที่ใช้ในการวิจัย 15

ประโยชน์ที่จะได้รับ 18

บทที่ 2 แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 20

ตอนที่ 1 มโนทัศน์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ 20

1.1 การทำหน้าที่ต่างกันของข้อสอบ (differential item functioning: DIF) 20

1.2 ประเภทการทำหน้าที่ต่างกันของข้อสอบ 23

1.3 การทำหน้าต่างกันของแบบสอบ (differential test functioning: DTF) 25

1.3.1 ตัวประมาณค่าการทำหน้าที่ต่างกันของข้อสอบทั่วไป(the generalized DIF estimators) 27

1.4 งานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ 28

ตอนที่ 2 มโนทัศน์การทำหน้าที่ต่างกันเป็นขั้น 35

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

หน้า

2.1 ที่มาของการทำหน้าที่ต่างกันเป็นขั้น (differential step functioning: DSF)..... 36
2.2 การอธิบายขั้น (step) ตามแนวคิดการทำหน้าที่ต่างกันเป็นขั้น..... 36
2.3 รูปแบบของผลการทำหน้าที่ต่างกันเป็นขั้น..... 40
2.4 เกณฑ์ในการกำหนดรูปแบบการทำหน้าที่ต่างกันเป็นขั้น..... 42
2.5 วิธีการเชิงสถิติสำหรับการตรวจสอบการทำหน้าที่ต่างกันเป็นขั้น..... 48
2.5.1 วิธีทฤษฎีตอบสนองข้อสอบ (IRT approach)..... 49
2.5.2 วิธีอัตราส่วนเลขออด (odds ratio approach)..... 49
2.5.3 วิธีถดถอยโลจิสติก (logistic regression approach)..... 50
2.6 การใช้ผลการทำหน้าที่ต่างกันเป็นขั้นเพื่อทดสอบการทำหน้าที่ต่างกันของข้อสอบ..... 50
2.7 งานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันเป็นขั้น..... 52
ตอนที่ 3 โครงการประเมินผลนักเรียนนานาชาติ (programme for international student assessment: PISA)..... 58
3.1 เครื่องมือที่ใช้ในโครงการประเมินผลนักเรียนนานาชาติ..... 59
3.1.1 การรู้เรื่องวิทยาศาสตร์ (scientific literacy)..... 59
3.1.2 การรู้เรื่องการอ่าน (reading literacy)..... 60
3.1.3 การรู้เรื่องคณิตศาสตร์ (mathematics literacy)..... 60
3.2 รูปแบบข้อสอบโครงการประเมินผลนักเรียนนานาชาติ..... 60
3.3 ข้อมูลสถานศึกษาจากการสำรวจในโครงการประเมินผลนักเรียนนานาชาติ..... 63
3.4 งานวิจัยที่เกี่ยวข้องโครงการประเมินผลนักเรียนนานาชาติ..... 72
ตอนที่ 4 แนวคิดเกี่ยวกับมูลค่าเพิ่มของการจัดการศึกษาและการวิเคราะห์พหุระดับ..... 76
4.1 มูลค่าเพิ่มของการจัดการศึกษา..... 76
4.2 ขั้นตอนและวิธีการที่ใช้วิเคราะห์มูลค่าเพิ่ม..... 79
4.3 การวิเคราะห์พหุระดับ..... 81

หน้า

4.4 งานวิจัยที่เกี่ยวข้องกับปัจจัยหรือตัวแปรที่ส่งผลต่อคุณภาพการจัดการศึกษาและการใช้โมเดลมูลค่าเพิ่มเพื่อประเมินคุณภาพการจัดการศึกษา 88
4.4.1 ปัจจัยหรือตัวแปรที่ส่งผลต่อคุณภาพการจัดการศึกษา 88
4.4.2 การใช้โมเดลมูลค่าเพิ่มเพื่อประเมินคุณภาพการจัดการศึกษา 90
ตอนที่ 5 กรอบแนวคิดในการวิจัย 101
บทที่ 3 วิธีดำเนินการวิจัย 104
ตอนที่ 1 ประชากรและกลุ่มตัวอย่าง 104
ตอนที่ 2 รูปแบบของแบบสอบประเมินการรู้เรื่องด้านวิทยาศาสตร์และรูปแบบข้อสอบ 106
ตอนที่ 3 คุณภาพเครื่องมือที่ใช้ในการวิจัย 107
ตอนที่ 4 ปัจจัยหรือตัวแปรที่นำมาใช้ในการวิจัย 108
ตอนที่ 5 การวิเคราะห์ข้อมูล 111
5.1 ตรวจสอบข้อสอบและแบบสอบที่นำมาศึกษา 112
5.2 วิเคราะห์ข้อมูลเบื้องต้นตามตัวแปรที่ศึกษา 112
5.3 ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและแบบสอบและการทำหน้าที่ต่างกันเป็นขั้น 113
5.4 ตรวจสอบเงื่อนไขการวิเคราะห์พหุระดับและความเป็นเอกมิติของแบบสอบและการวิเคราะห์พหุระดับใช้ใช้โมเดลมูลค่าเพิ่มพหุระดับ 116
5.5 เปรียบเทียบประสิทธิภาพโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ 119
บทที่ 4 ผลการวิเคราะห์ข้อมูล 121
ตอนที่ 1 ผลการวิเคราะห์ข้อมูลพื้นฐาน 122
ตอนที่ 2 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ และการทำหน้าที่ต่างกันเป็นขั้น 130
2.1 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ และการทำหน้าที่ต่างกันเป็นขั้นตามตัวแปรที่ศึกษา 130

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

หน้า

2.2 ผลการพิจารณาการตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันและโครงสร้างข้อสอบหลังการตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน 136

2.3 ผลการประมาณค่าคะแนนความสามารถของนักเรียนหลังจากตรวจสอบโครงสร้างข้อสอบ..... 151

ตอนที่ 3 ผลการตรวจสอบเงื่อนไขการวิเคราะห์พหุระดับและความเป็นเอกมิติของแบบสอบ ... 154

3.1 การตรวจสอบเงื่อนไขการวิเคราะห์พหุระดับ 154

3.1.1 การตรวจสอบภาวะร่วมเส้นตรงพหุ..... 155

ตอนที่ 4 ผลการเปรียบเทียบประสิทธิภาพโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ 161

4.1 ลักษณะโมเดลการประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ 161

4.2 การวิเคราะห์โมเดลมูลค่าเพิ่มพหุระดับ..... 165

4.3 การสร้างเกณฑ์การตัดสินการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม 171

4.4 วิเคราะห์และเปรียบเทียบความสัมพันธ์ระหว่างโมเดลมูลค่าเพิ่มพหุระดับของ 2 โมเดล..... 173

4.5 เปรียบเทียบความสอดคล้องการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่มและการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่มของ 2 โมเดล 177

บทที่ 5 สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ 190

5.1 สรุปผลการวิจัย..... 191

ประเด็นที่ 1 ผลจากการวิเคราะห์ข้อมูลพื้นฐานในการวิจัย 191

ประเด็นที่ 2 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและของแบบสอบเมื่อใช้โมเดลมูลค่าเพิ่มพหุระดับที่ต่างกัน 193

2.1 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 193

2.2 ผลการตรวจสอบการทำหน้าที่ต่างกันของแบบสอบ 194

2.3 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและแบบสอบภายหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน 195

หน้า

ประเด็นที่ 3 ผลการเปรียบเทียบประสิทธิภาพโมเดลมูลค่าเพิ่มการจัดการศึกษาวิทยาศาสตร์ระหว่าง 2 โมเดล 197
3.1 ผลการตรวจสอบเงื่อนไขการวิเคราะห์พหุระดับเบื้องต้น 197
3.2 ผลการเปรียบเทียบประสิทธิภาพโมเดลการประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ 199
3.3 ผลการศึกษาความสัมพันธ์ระหว่างโมเดลมูลค่าเพิ่มพหุระดับของ 2 โมเดล 201
3.4 ผลการเปรียบเทียบความสอดคล้องของประสิทธิผลของโรงเรียน การจัดอันดับคุณภาพภาพและการจัดกลุ่มคุณภาพระหว่างโมเดลมูลค่าเพิ่มพหุระดับระหว่าง 2 โมเดล 202
5.2 อภิปรายผลการวิจัย 205
5.2.1 ข้อสอบที่เกิดจากการทำหน้าที่ต่างกันของข้อสอบและแบบสอบตามโครงการประเมินผลนักเรียนนานาชาติ ปี 2009 206
5.2.1.1 การศึกษาการทำหน้าที่ต่างกันของข้อสอบตามโครงการประเมินผลนักเรียนนานาชาติ ปี 2009 206
5.2.1.2 การศึกษาการทำหน้าที่ต่างกันของแบบสอบตามโครงการประเมินผลนักเรียนนานาชาติ ปี 2009 211
5.2.1.3 ความสัมพันธ์ระหว่างการทำหน้าที่ต่างกันของข้อสอบและแบบสอบตามโครงการประเมินผลนักเรียนนานาชาติ ปี 2009 213
5.2.2 ผลการเปรียบเทียบประสิทธิภาพโมเดลมูลค่าเพิ่มการจัดการศึกษาวิทยาศาสตร์ 214
5.2.2.1 ความสำคัญของการนำโมเดลมูลค่ามาใช้ในการพัฒนาคุณภาพการจัดการศึกษาวิทยาศาสตร์ 215
5.2.2.2 การคัดเลือกตัวแปรเพื่อนำมาศึกษาประสิทธิภาพโมเดลมูลค่าเพิ่มการจัดการศึกษาวิทยาศาสตร์ 216
5.2.2.3 ผลการเปรียบเทียบประสิทธิภาพโมเดลมูลค่าเพิ่มพหุระดับเพื่อการประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ 217
5.2.2.4 ความสัมพันธ์ระหว่างโมเดลมูลค่าเพิ่มพหุระดับของ 2 โมเดล 221

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

หน้า

5.2.2.5 เปรียบเทียบความสอดคล้องของประสิทธิผลของโรงเรียนการจัดอันดับคุณภาพและการจัดกลุ่มคุณภาพระหว่างโมเดลมูลค่าเพิ่มพหุระดับระหว่าง 2 โมเดล 222

5.3 ข้อเสนอแนะ 226

5.3.1 ข้อเสนอแนะในการนำผลการวิจัยไปใช้ประโยชน์ 226

5.3.2 ข้อเสนอแนะสำหรับการวิจัยครั้งต่อไป 227

รายการอ้างอิง 229

ภาคผนวก ก ข้อมูลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ และการทำหน้าที่ต่างกันเป็นขั้น 242

ภาคผนวก ข ข้อสอบที่ถูกตัดแยกรายฉบับและภาพรวม 252

ภาคผนวก ค ผลการวิเคราะห์โมเดลมูลค่าเพิ่มพหุระดับ 4 โมเดล 256

ภาคผนวก ง คะแนนเฉลี่ยการประเมินการรู้เรื่องวิทยาศาสตร์ คะแนนมูลค่าเพิ่ม การจัดอันดับคะแนนมูลค่าเพิ่ม และการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่มระหว่าง 2 โมเดล 269

ประวัติผู้เขียนวิทยานิพนธ์ 276

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญตาราง

ตารางที่ 1 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 23
ตารางที่ 2 สรุปผลองค์ความรู้จากงานวิจัยการทำหน้าที่ต่างกันของข้อสอบ 33
ตารางที่ 3 การจัดกลุ่มภายในรูปแบบการทำหน้าที่ต่างกันเป็นขั้น 40
ตารางที่ 4 ผลการทำหน้าที่ต่างกันเป็นขั้น 41
ตารางที่ 5 ลักษณะการทำหน้าที่ต่างกันเป็นขั้นตามหมวดหมู่ (a taxonomy of DSF) 41
ตารางที่ 6 สัญลักษณ์และเกณฑ์การแปลความหมายของ $\hat{\lambda}_j$ 44
ตารางที่ 7 รูปแบบการทำหน้าที่ต่างกันตามการจัดกลุ่มตามผลการทำหน้าที่ต่างกันเป็นขั้น (a DSF effect categorization scheme) 45
ตารางที่ 8 สรุป net DIF และ global DIF 52
ตารางที่ 9 สรุปผลองค์ความรู้จากงานวิจัยการทำหน้าที่ต่างกันเป็นขั้น 55
ตารางที่ 10 รอบการประเมินโครงการประเมินผลนักเรียนนานาชาติ 59
ตารางที่ 11 สเกลการวัดและความหมายของตัวแปร 64
ตารางที่ 12 สรุปผลองค์ความรู้จากงานวิจัยโครงการประเมินผลนักเรียนนานาชาติ 74
ตารางที่ 13 สรุปผลองค์ความรู้จากงานวิจัยโมเดลมูลค่าเพิ่ม 96
ตารางที่ 14 ผลการสังเคราะห์ตัวแปรทำนายที่มีอิทธิพลต่อผลสัมฤทธิ์ทางการเรียน 99
ตารางที่ 15 อันดับปัจจัยที่มีผลต่อคุณภาพการจัดการศึกษาที่สอดคล้องกับโครงการประเมินผลนักเรียนนานาชาติ 2009 100
ตารางที่ 16 กรอบการสุ่มกลุ่มตัวอย่างจำแนกตามภาคพื้นที่ภูมิศาสตร์และสังกัดโรงเรียน 104
ตารางที่ 17 จำนวนโรงเรียนกลุ่มตัวอย่าง PISA 2009(สังกัดโรงเรียน x พื้นที่ภูมิศาสตร์) 105
ตารางที่ 18 จำนวนกลุ่มตัวอย่างนักเรียนในโครงการประเมินผลนานาชาติจำแนกตามสังกัด 105
ตารางที่ 19 ลักษณะข้อสอบที่ใช้ในจำแนกตามสมรรถนะและความรู้ทางวิทยาศาสตร์ 106
ตารางที่ 20 น้ำหนักองค์ประกอบและความเที่ยง 108

ตารางที่ 21 องค์ประกอบของ ESCS ในปี 2000, 2003, 2006, 2009 108
ตารางที่ 22 สเกลการวัดและความหมายตัวแปรที่ใช้ในการศึกษา 109
ตารางที่ 23 สัญลักษณ์และความหมายตัวแปรที่ใช้ในการศึกษาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ 111
ตารางที่ 24 จำนวนนักเรียนก่อนการคัดเลือกและที่ใช้ในการวิจัย 112
ตารางที่ 25 อิทธิพลการทำหน้าที่ต่างกันของข้อสอบที่ใช้ประเมินแบบสอบที่ตรวจพบการทำหน้าที่ต่างกัน 113
ตารางที่ 26 ขนาดอิทธิพลความแปรปรวนการทำหน้าที่ต่างกันเป็นขั้น 114
ตารางที่ 27 สมการวิเคราะห์มูลค่าเพิ่มพหุระดับตามโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ของสถานศึกษาทั้ง 2 โมเดล 118
ตารางที่ 28 ช่วงคะแนนการตัดสินการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม (VAS rating) 119
ตารางที่ 29 ช่วงการตัดสินความสอดคล้องค่าสัมประสิทธิ์ของ kappa 120
ตารางที่ 30 ความถี่และร้อยละของตัวแปรจัดประเภทจำแนกตามตัวแปรระดับนักเรียนและสถานศึกษา 123
ตารางที่ 31 ค่าสถิติพื้นฐานของตัวแปรต่อเนื่องจำแนกตามตัวแปรระดับนักเรียนและสถานศึกษา 125
ตารางที่ 32 ค่าสถิติพื้นฐานคะแนนความสามารถทางวิทยาศาสตร์จำแนกตามตัวแปรระดับนักเรียนและสถานศึกษา 128
ตารางที่ 33 ผลข้อสอบที่ตรวจพบทำหน้าที่ต่างกันรวมทั้ง 5 ตัวแปรในแบบสอบทั้ง 9 ฉบับแยกตามสมรรถนะทางวิทยาศาสตร์ 132
ตารางที่ 34 ข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันในแต่ละตัวแปรที่ศึกษา 134
ตารางที่ 35 ผลการตรวจสอบการทำหน้าที่ต่างกันในแบบสอบ 9 ฉบับตามตัวแปรที่ศึกษา 135
ตารางที่ 36 ลักษณะข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันทซึ่งถูกตัดตามตัวแปรที่ศึกษา 137
ตารางที่ 37 การเปรียบเทียบร้อยละข้อสอบที่ถูกตัด ค่าความเที่ยง และช่วงพารามิเตอร์ของข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันทั้ง 9 ฉบับ 138
ตารางที่ 38 ผลการตรวจสอบแบบสอบที่ตรวจพบการทำหน้าที่ต่างกันตามตัวแปรที่ศึกษา 139
ตารางที่ 39 สัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรตามสมรรถนะทางวิทยาศาสตร์ 143

ตารางที่ 40 สรุปดัชนีที่ใช้ในการตรวจสอบความสอดคล้องกลมกลืนของโมเดลกับข้อมูลเชิงประจักษ์โมเดลก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน 146

ตารางที่ 41 สรุปดัชนีที่ใช้ในการตรวจสอบความสอดคล้องกลมกลืนของโมเดลกับข้อมูลเชิงประจักษ์ของโมเดลหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน..... 147

ตารางที่ 42 ค่าสถิตพื้นฐานคะแนนความสามารถของนักเรียนก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันในแบบสอบ 9 ฉบับ 151

ตารางที่ 43 ค่าสถิตพื้นฐานคะแนนความสามารถของนักเรียนหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันในแบบสอบ 9 ฉบับ..... 152

ตารางที่ 44 สรุปผลการแปลงค่าคะแนนสเกลการประเมินผลการรู้เรื่องวิทยาศาสตร์ 153

ตารางที่ 45 เปรียบเทียบคะแนนเฉลี่ยของตัวแปรก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน..... 154

ตารางที่ 46 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรระดับนักเรียนก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน..... 156

ตารางที่ 47 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรระดับนักเรียนหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน..... 157

ตารางที่ 48 ค่า Tolerance และ VIF จากผลการทดสอบความสัมพันธ์ระหว่างตัวแปรทำนายระดับนักเรียนก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน 158

ตารางที่ 49 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรระดับสถานศึกษาก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน 159

ตารางที่ 50 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรระดับสถานศึกษาหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน 160

ตารางที่ 51 ค่า Tolerance และ VIF จากผลการทดสอบความสัมพันธ์ระหว่างตัวแปรทำนายระดับสถานศึกษาก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน..... 161

ตารางที่ 52 สรุปค่าสัมประสิทธิ์ของตัวแปรระดับนักเรียนและสถานศึกษาทั้ง 4 โมเดล..... 170

ตารางที่ 53 ค่าสถิติพื้นฐานของคะแนนมูลค่าเพิ่มพหุระดับจากการประเมินผลการรู้เรื่องวิทยาศาสตร์ของสถานศึกษาทั้ง 2 โมเดล..... 172

ตารางที่ 54 ช่วงคะแนนการตัดสินการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม 173

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 55 ความสัมพันธ์ระหว่าง SCIEmean, VAS, VAS ranking และ VAS rating 176

ตารางที่ 56 ความสัมพันธ์ระหว่างคะแนนจากการประมาณค่าความสามารถและคะแนน
มูลค่าเพิ่มระหว่าง 2 โมเดล 177

ตารางที่ 57 ผลการเปรียบเทียบความสอดคล้องประสิทธิผลคะแนนมูลค่าเพิ่มของสถานศึกษา 178

ตารางที่ 58 ผลความสอดคล้องการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่มโดยค่าสถิติทดสอบ
Wilcoxon Signed Ranks Test 179

ตารางที่ 59 ผลความสอดคล้องการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม (VAS rating) โดยค่า
สัมประสิทธิ์ของ Kappa 180

ตารางที่ 60 เปรียบเทียบความสอดคล้องตามตัวแปรสถานศึกษา 184

ตารางที่ 61 สรุปสหสัมพันธ์ภายในชั้นและสัมประสิทธิ์การทำนาย 200

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญรูปภาพ

รูปภาพที่ 1 แสดงลักษณะการทำหน้าที่ต่างกันของข้อสอบ (DIF) 25
รูปภาพที่ 2 แสดงการทำหน้าที่เป็นขั้นในขั้นที่ 1 37
รูปภาพที่ 3 แสดงการทำหน้าที่เป็นขั้นในขั้นที่ 2 38
รูปภาพที่ 4 แสดงการทำหน้าที่เป็นขั้นในขั้นที่ 3 38
รูปภาพที่ 5 แสดงการทำหน้าที่เป็นขั้นในข้อสอบที่มีการตรวจให้คะแนน 4 ค่า มาจาก Penfield, Gattamorta, และ Childs (2009) 39
รูปภาพที่ 6 แสดงการทำหน้าที่เป็นขั้นในข้อสอบที่มีการตรวจให้คะแนน 4 ค่า ในกลุ่มอ้างอิง (เส้นทึบ) และกลุ่มเปรียบเทียบ(เส้นปะ) มาจาก Penfield, Gattamorta, และ Childs, 2009) 39
รูปภาพที่ 7 แสดงผลการทำหน้าที่ต่างกันเป็นขั้นในข้อสอบที่มีการตรวจให้คะแนนพหุวิภาค 4 ค่า พร้อมการเกิด 6 ลักษณะ มาจาก Wood (2011) 48
รูปภาพที่ 8 ตัวอย่างข้อสอบประเภทเลือกตอบและการให้คะแนน มาจากสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (2551) 61
รูปภาพที่ 9 ตัวอย่างข้อสอบเลือกตอบเชิงซ้อนและการให้คะแนน มาจากสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (2551) 62
รูปภาพที่ 10 ตัวอย่างข้อสอบแบบตอบอิสระแบบปิดและการให้คะแนน มาจากสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (2551) 62
รูปภาพที่ 11 ตัวอย่างข้อสอบแบบตอบอิสระแบบเปิดและการให้คะแนน มาจากสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (2551) 63
รูปภาพที่ 12 แสดงโมเดลวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลก่อนและหลังตัดข้อสอบ DIF ในแบบสอบ 9 ฉบับและรวมทุกฉบับ 150
รูปภาพที่ 13 ฮีสโตแกรม (Histrogram) คะแนนมูลค่าเพิ่มจากโมเดลการประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์จำนวน 223 แห่ง 172

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

บทที่ 1
บทนำ

ความเป็นมาและความสำคัญของปัญหา

คุณภาพของแบบทดสอบเป็นสิ่งที่มีความสำคัญในเรื่องการวัดและประเมินผลทางด้านการศึกษา การที่จะพัฒนาแบบทดสอบให้มีคุณภาพนั้นจะต้องคำนึงถึงคุณภาพแบบทดสอบ ซึ่งแบบทดสอบที่มีคุณภาพนั้นจำเป็นอย่างยิ่งที่จะต้องประกอบไปด้วยความเที่ยง (reliability) และความตรง (validity) ที่ได้มาตรฐาน โดยเฉพาะความตรงนั้นนับว่าเป็นหัวใจที่สำคัญยิ่งของแบบทดสอบที่แสดงถึงความสามารถในการวัดได้ถูกต้องแม่นยำ ถ้าสามารถวัดได้ค่าที่ใกล้เคียงกับค่าคุณลักษณะที่แท้จริงเพียงใดถือว่าการวัดนั้นมีความตรงมากขึ้นเท่านั้น (Ayala, 2008; Demar, 2010; Embretson & Reise, 2000; Holland & Wainer, 1993; Kunnan, 2000; ศิริชัย กาญจนวาสี, 2550ก) โดยทั่วไปการตรวจสอบความตรงของแบบทดสอบ แบ่งออกได้เป็น 3 ประเภทหลัก คือ ความตรงตามเนื้อหา (content validity) ความตรงตามเกณฑ์สัมพันธ์ (criterion-related validity) และความตรงตามโครงสร้าง (construct validity) นอกจากนี้คุณสมบัติที่สำคัญอีกประการหนึ่ง คือ ความยุติธรรมของการทดสอบซึ่งผ่านการศึกษาการตรวจสอบคุณภาพความตรงของแบบทดสอบ แต่เดิมใช้คำว่า การตรวจสอบ “ความลำเอียงของข้อสอบหรือแบบทดสอบ (item/test bias) ซึ่งความหมายเชิงลบ อีกทั้งการตัดสินในยุคก่อนว่าข้อสอบหรือแบบทดสอบมีความลำเอียงหรือไม่มักไม่ได้คำนึงถึงวิธีการทางสถิติทำให้เกิดความไม่ชัดเจนในเกณฑ์ที่ใช้ตัดสิน ต่อมาเมื่อวิทยาการมีการพัฒนามากยิ่งขึ้นจึงได้นำวิธีการทางสถิติเข้ามาเป็นเกณฑ์ร่วมในการตัดสินและเปลี่ยนแปลงมาใช้คำว่า “การทำหน้าที่ต่างกันของข้อสอบหรือแบบทดสอบ” จนถึงปัจจุบัน (differential item/test functioning; DIF/DTF) (Holland & Wainer, 1993; Shealy & Stout, 1993; ศิริชัย กาญจนวาสี, 2550ก)

นักวิชาการทั้งในและต่างประเทศได้ให้ความหมายการทำหน้าที่ต่างกันของข้อสอบไว้หลากหลาย ศิริชัย กาญจนวาสี (2550ก) ได้กล่าวไว้ว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึงการที่ข้อสอบทำให้ผู้สอบจากต่างกลุ่มกันที่มีความสามารถหรือคุณลักษณะที่มุ่งวัดเท่ากัน มีโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกันหรือมีฟังก์ชันการตอบสนองข้อสอบแตกต่างกัน เช่นเดียวกับ Holland และ Thayer (1993) ได้กล่าวไว้ว่า ผู้สอบสองกลุ่มมีโอกาสในการตอบข้อสอบถูกซึ่งเกี่ยวข้องกับระดับความสามารถที่เปรียบเทียบกันได้ นอกจากนี้ในแง่ของการทดสอบทางการศึกษาและจิตวิทยา การทำหน้าที่ต่างกันของข้อสอบ ยังหมายถึง โอกาสในการตอบข้อสอบได้ถูกต้อง

ระหว่างความสามารถของผู้เข้าสอบที่เท่ากันซึ่งผู้สอบอาจจะมีความแตกต่างกันตามเชื้อชาติ เพศ หรือกลุ่มย่อยอื่นๆ ก็เป็นได้ (Westers & Kelderman, 1992)

นอกจากนี้ Park และ Lautenschlager (1990 อ้างถึงใน กาญจนา วัธนสุนทร, 2537) ได้ให้ข้อสังเกตเกี่ยวกับความลำเอียงหรือการทำหน้าที่ต่างกันของข้อสอบ อยู่ 2 ประเด็น คือ 1) การตอบข้อสอบมีอิทธิพลจากแหล่งความแปรปรวนอื่นๆ นอกเหนือจากความแตกต่างบนมิติที่สนใจ 2 แหล่ง คือ แหล่งความแปรปรวนจากข้อสอบและแหล่งความแปรปรวนจากผู้สอบ และ 2) แหล่งความแปรปรวนภายนอกมีผลต่อการตอบข้อสอบในลักษณะที่ทำให้กลุ่มผู้สอบย่อยบางกลุ่มแตกต่างจากกลุ่มอื่นอย่างเป็นระบบทำให้เกิดการไม่ยุติธรรม ส่วนการวิจัยของ Adam (2003) และ Adam และคณะ (2007) ได้ระบุความคลาดเคลื่อนที่เกิดจากรูปแบบข้อสอบหรือการจัดพิมพ์ที่ผิดพลาดของข้อสอบฉบับย่อยทำให้เกิดข้อสอบตรวจพบการทำหน้าที่ต่างกันได้ ส่วน Le (2009) พบว่า ข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันส่วนหนึ่งมาจากรูปแบบข้อสอบและเนื้อหาที่ใช้ออกข้อสอบ และการศึกษาของ Wetzel และ Carstensen (2013) พบว่า คะแนนความสามารถอาจจะเอนเอียงเข้าข้างยังด้านใดด้านหนึ่งขึ้นอยู่กับว่าข้อสอบข้อนั้นได้รับอิทธิพลจากความยากตามเนื้อหาของข้อสอบ

ทั้งนี้การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสามารถแบ่งมิติในการตรวจสอบออกเป็น 2 ประเภท คือ การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในข้อสอบแบบตรวจให้คะแนน 2 ค่าหรือทวิวิภาค (Dichotomous items) และแบบตรวจให้คะแนนหลายค่าหรือพหุวิภาค (Polytomous items) การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในข้อสอบแบบตรวจให้คะแนน 2 ค่า มาจากแนวคิดทฤษฎีการตอบสนองข้อสอบ (IRT) ซึ่งสามารถอธิบายด้วยฟังก์ชันลักษณะข้อสอบหรือโค้งลักษณะข้อสอบ (ICCs) ส่วนการตรวจให้คะแนนหลายค่าเป็นโมเดลความสัมพันธ์ที่ไม่ใช่เชิงเส้นตรงระหว่างความสามารถของผู้สอบกับโอกาสของการเลือกตอบแต่ละรายการคำตอบที่กำหนดให้ ซึ่งในเครื่องมือมักมีหลายคำตอบให้ผู้เข้าสอบได้เลือก (Li & Stout, 1996; Swaminathan & Rogers, 1990; ศิริชัย กาญจนวาสี, 2550ก) นอกจากนี้การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบยังแบ่งได้เป็น 2 กลุ่ม คือ กลุ่มที่หนึ่งตามทฤษฎีการทดสอบแบบดั้งเดิม ซึ่งวิธีดังกล่าวมีจุดอ่อนตรงข้อตกลงเบื้องต้น คือ คะแนนความคลาดเคลื่อน ค่าพารามิเตอร์ข้อสอบและแบบทดสอบผันแปรไปตามกลุ่มข้อสอบ และกลุ่มที่สองเกิดขึ้นภายหลัง โดยการประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบเพื่อผ่อนคลายข้อตกลงเบื้องต้นและแก้ไขจุดอ่อนบางประการจากทฤษฎีการทดสอบแบบดั้งเดิม ศิริชัย กาญจนวาสี (2550ก) จำแนกเป็น 2 กลุ่ม คือ 1) ใช้การเปรียบเทียบค่าประมาณพารามิเตอร์ของข้อสอบระหว่างกลุ่มผู้สอบ มีหลักแนวคิดคือใช้การเปรียบเทียบค่าประมาณพารามิเตอร์ของข้อสอบระหว่างกลุ่มอ้างอิง (reference: R) และกลุ่มเปรียบเทียบ (focal: F) โดยการทดสอบนัยสำคัญทางสถิติ และ 2) ใช้การเปรียบเทียบค่าประมาณ

ฟังก์ชันการตอบสนองข้อสอบระหว่างกลุ่มผู้สอบ โดยการวัดพื้นที่ระหว่างฟังก์ชันการตอบข้อสอบ มีหลักแนวคิด คือ ใช้การวัดพื้นที่ของฟังก์ชันการตอบสนองข้อสอบ (item response function: IRFs) ระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบโดยตรง แล้วตัดสินการทำหน้าที่ต่างกันของข้อสอบโดยนำดัชนีที่คำนวณได้เปรียบเทียบกับเกณฑ์ที่กำหนดขึ้นหรือใช้การทดสอบนัยสำคัญทางสถิติ นอกจากการแบ่งตามทฤษฎีแล้วการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบยังแบ่งตามสถิติที่ใช้ตรวจสอบโดยจำแนกได้เป็น 2 กลุ่ม คือ กลุ่มสถิตพาราเมทริก (parametric approach) เช่น การวิเคราะห์ความแปรปรวน (ANOVA) วิธีถดถอยโลจิสติก (logistic regression) วิธีวัดพื้นที่ความแตกต่างระหว่างโค้งการตอบสนองข้อสอบ (IRT-D^2) วิธีไค-สแควร์ของลอร์ด (lord's χ^2) วิธีอัตราส่วนไลค์ลิฮูดทั่วไป (general IRT likelihood ratio) วิธีอัตราส่วนไลค์ลิฮูดลอกลิเนียร์ (loglinear IRT-LR likelihood ratio) การวิเคราะห์ถดถอยโลจิสติกพหุวิภาค (polytomous logistic regression) วิธีอัตราส่วนไลค์ลิฮูดในรูปทั่วไป (general IRT likelihood ratio) การให้คะแนนบางส่วน (partial credit model) และกลุ่มสถิตินันพาราเมทริกซึ่งไม่มีข้อตกลงเหมือนกับสถิติพาราเมทริก เช่น วิธีแปลงค่าความยากของข้อสอบ (transformed item difficulty: TID) วิธีแมนเทล-แฮนส์เซล (mantel-haenszel: MH) วิธีดัชนีมาตรฐานการปรับให้เป็นมาตรฐานด้วยน้ำหนักตัวประกอบ (standardization: STND) วิธีซิปเทสท์ (SIBTEST) วิธีดัชนีมาตรฐานพหุวิภาค (polytomous STND) วิธีแมนเทล-แฮนส์เซลทั่วไป (general mantel-haenszel: GMH) วิธีซิปเทสท์พหุวิภาค (polytomous SIBTEST) วิธีการให้คะแนนบางส่วนทั่วไป (generalized partial credit model) และอีกวิธีที่ถูกนำเสนอโดย Penfield (2007b, 2008, 2010a) คือ วิธี การทำหน้าที่ต่างกันเป็นขั้น (differential step functioning: DSF)

Penfield (2007b, 2008, 2010a) ได้เสนอการทำหน้าที่ต่างกันเป็นขั้น ซึ่งเป็นวิธีตรวจสอบข้อสอบแบบนันพาราเมทริกใช้การตรวจสอบความแตกต่างระหว่างกลุ่ม (between-group differences) ในแต่ละขั้น (step) ของข้อสอบที่มีการตรวจให้คะแนนหลายค่า แต่ละขั้นแสดงถึงการมีโอกาสได้เปรียบ (probability of advancing) หรือ stepping และแต่ละขั้นอาจเกิดเหตุการณ์ที่กลุ่มอ้างอิงได้เปรียบกลุ่มเปรียบเทียบหรือกลุ่มเปรียบเทียบได้เปรียบกลุ่มอ้างอิงก็ได้ ส่งผลให้การทำหน้าที่ต่างกันเป็นขั้นครอบคลุมการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (Penfield, Alvarez, & Lee, 2008; Penfield, Gattamorta, & Childs, 2009)

แนวคิดการทำหน้าที่ต่างกันเป็นขั้นมาจากการรวมผลการทำหน้าที่ต่างกันของข้อสอบในทุกขั้นของข้อสอบที่มีการตรวจให้คะแนนหลายค่า ถ้าผลการทำหน้าที่ต่างกันเป็นขั้น (DSF effects) เพิ่มขึ้นผลการทำหน้าที่ต่างกันของข้อสอบมีขนาดเพิ่มขึ้นด้วย ข้อสอบที่มีการตรวจให้คะแนนหลายค่ามีกลุ่มตอบสนองอยู่ r ลำดับ สามารถกำหนดการทำหน้าที่เป็นขั้นได้ทั้งหมด J=r-1 ขั้น โดยแต่ละขั้นจะแทนด้วยสัญลักษณ์ j ซึ่ง j มีค่าตั้งแต่ j = 1, 2, ... , J (Penfield, 2007b; Penfield,

Alvarez, & Lee, 2008; Penfield, Gattamorta, & Childs, 2009) นอกจากนี้การวัดตามวิธีการทำหน้าที่ต่างกันเป็นขั้นสามารถแปรเปลี่ยนทั้งขนาด (magnitude) และ/หรือ สัญลักษณ์ (sign) ข้ามขั้นภายใต้ J ขั้นของข้อสอบที่มีการตรวจให้คะแนนหลายค่า ทำให้ Penfield (2010b) ทำการศึกษาการทำหน้าที่ต่างกันแบบตัดกันของข้อสอบ (crossing DIF) ใน graded response model (GRM) และ partial credit model (PCM) เกิดรูปแบบการทำหน้าที่ต่างกันเป็นขั้น 12 เงื่อนไข ถ้ามีลักษณะลู่เข้า (convergent) หมายถึง ผลของการทำหน้าที่ต่างกันเป็นขั้นมีเครื่องหมายเดียวกัน (sign) แต่มีขนาด (magnitude) ที่แตกต่างกัน และลักษณะลู่ออก (divergent) หมายถึง เครื่องหมายมีลักษณะตรงกันข้ามหรือเปลี่ยนเครื่องหมายข้ามขั้น นอกจากนี้การทำหน้าที่ต่างกันเป็นขั้นสามารถกำหนดขนาดตามหมวดหมู่ของ ETS (the educational testing service) โดยใช้หลักการอัตราส่วนลอกเลขออดร่วมเป็นระดับขั้น (the step-level common log-odds ratio) แบ่งออกเป็น ขนาดเล็ก (small: S) ขนาดกลาง (medium: M) และขนาดใหญ่ (large: L) ตามการศึกษาของ Penfield, Gattamorta, และ Childs (2009) ได้วิเคราะห์ผลการทำหน้าที่ต่างกันเป็นขั้นกับข้อสอบวิชาคณิตศาสตร์ที่มีการตรวจให้คะแนนแบบหลายค่า จำนวน 30 ข้อ โดยใช้ตัวประมาณค่าอัตราส่วนลอกเลขออดร่วมเป็นระดับขั้นสะสม (the cumulative step-level log-odds ratio estimator) ร่วมกับการทดสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้สถิติทดสอบอัตราส่วนแต้มต่อลัว-อเกรสตีสะสม (the Liu-Agresti cumulative common log-odds ratio: LA) เพื่อระบุขนาดการทำหน้าที่ต่างกันเป็นขั้นในแต่ละขั้นของข้อสอบ

เมื่อมีการตรวจพบว่าข้อสอบหรือแบบทดสอบข้อใดที่มีการทำหน้าที่ต่างกันแล้วก็อาจจะตัดข้อสอบหรือนำข้อสอบข้อนั้นไปปรับปรุงใหม่ (Westers & Kelderman, 1992) แต่การตัดข้อสอบข้อนั้นออกไปอาจกระทบต่อความตรงเชิงเนื้อหา เนื่องจากข้อสอบที่ลำเอียงไม่จำเป็นต้องเป็นข้อสอบที่ไม่ดีไปทั้งหมดแต่อาจมีคุณลักษณะที่ดีของข้อสอบอย่างอื่น เช่น เป็นข้อสอบที่ให้สารสนเทศที่ดีที่สุดได้ (most information) หรือเป็นตัวทำนายที่ดีที่สุดได้ (best predictor) ก็เป็นได้ (Shealy & Stout, 1993) การพิจารณาข้อสอบบางข้อโดยการนำกลับไปรวมกับแบบทดสอบฉบับนั้นเช่นเดิมก็เพื่อให้แบบทดสอบฉบับนั้นเกิดความยุติธรรมกับกลุ่มผู้สอบย่อยทุกกลุ่มเท่ากัน วิธีการดังกล่าวก็คือการศึกษาการทำหน้าที่ต่างกันของแบบสอบ (differential test functioning) ซึ่งเป็นกระบวนการตรวจสอบความแตกต่างภายใต้โค้งลักษณะแบบสอบ (TCCs) และเป็นวิธีการที่สำคัญจนกล่าวได้ว่า การทดสอบการทำหน้าที่ต่างกันในข้อสอบในข้อใดข้อหนึ่งอาจจะมีความสำคัญแต่อาจจะไม่มากนักที่จะมีผลกระทบต่อผลที่ได้จากแบบสอบทั้งฉบับ (Jong, 2012) ดังนั้นการพิจารณาดัชนีการวิเคราะห์การทำหน้าที่ต่างกันของแบบทดสอบทั้งฉบับจึงมีประโยชน์เพื่อใช้พิจารณาแบบสอบที่เข้าข้างกลุ่มใดกลุ่มหนึ่ง ซึ่งจะมีประโยชน์ในการใช้เป็นข้อสารสนเทศในการตัดสินประสิทธิภาพของเครื่องมือ การจัดลำดับ หรือการคัดเลือก

วิธีการเชิงสถิติที่ใช้ตรวจสอบการทำหน้าที่ต่างกันของแบบทดสอบมีความแตกต่างจากการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 3 ประการ ได้แก่ ประการแรก คือ การทำหน้าที่ต่างกันของแบบทดสอบอาจจะรวมความแตกต่างของฟังก์ชันการตอบสนองของแบบทดสอบสองชุดหรือแบบทดสอบย่อย ประการที่สอง คือ การทำหน้าที่ต่างกันของแบบทดสอบรวมความแตกต่างของการทำหน้าที่ต่างกันของข้อสอบทุกๆ ข้อ แม้ว่าดัชนีการทำหน้าที่ต่างกันของข้อสอบจะเข้าใกล้ศูนย์สำหรับการรวมของข้อสอบทุกข้อในแบบทดสอบก็ตาม และประการที่สาม คือ การทำหน้าที่ต่างกันของแบบทดสอบเป็นความแตกต่างระหว่างกลุ่มในความแปรปรวนร่วมของข้อสอบและโครงสร้างเฉลี่ยในวิธีการวิเคราะห์องค์ประกอบ ซึ่งผลการวิเคราะห์ประกอบด้วยน้ำหนักข้อสอบนั้นมักถูกพิจารณาถัดมาในแง่ของค่าเฉลี่ย ความแปรปรวน และความแปรปรวนร่วม (Ackerman 1992, 1996; Camilli, 1992; Cammilli & Penfield, 1997; Longford, 1993, 1995; Muthen, 1985; Raju, Van der Linden & Fleer, 1995; Shealy & Stout, 1993; Zwick, 1995 cited in Camilli & Congdon, 1999)

โมเดลการวัดการทำหน้าที่ต่างกันของแบบทดสอบจำแนกได้เป็น 3 รูปแบบ คือ โมเดลการทำหน้าที่ต่างกันของแบบทดสอบของข้อสอบที่มีการให้คะแนนสองค่า ซึ่งจำแนกได้เป็น 3 แนวคิด ได้แก่ แนวคิดที่หนึ่งมาจาก Shealy และ Stout (1993) เสนอกระบวนการตรวจสอบแบบ SIBTEST แนวคิดที่สองมาจากกระบวนการ DFITS ซึ่งใช้การประมาณค่าความแตกต่างกำลังสองระหว่างกลุ่มที่ถูกคาดหวัง (the expected between-group squared difference) ของคะแนนจริง และแนวคิดที่สามมาจากโมเดลอิทธิพลเชิงสุ่ม (a random effects model) และต่อมาประเมินความแปรปรวนในผลการทำหน้าที่ต่างกันของข้อสอบในทุกข้อของแบบทดสอบ นอกจากนี้โมเดลตรวจสอบการทำหน้าที่ต่างกันของแบบทดสอบในข้อสอบที่มีการตรวจให้คะแนนหลายค่าสามารถใช้วิธี SIBTEST และ DFITS เข้ามาช่วยในการตรวจสอบ และอีกแนวคิดที่ถูกนำเสนอเป็นโมเดลการตรวจให้คะแนนทั้งสองค่าและหลายค่าหรือแบบผสมผสาน (mixed format tests) ในแบบสอบฉบับเดียวกัน Penfield และ Algina (2006) ได้ศึกษาตัวประมาณค่าการทำหน้าที่ต่างกันของข้อสอบแบบทั่วไป (the generalized DIF estimators) ทั้งประเภทถ่วงน้ำหนักและไม่ถ่วงน้ำหนัก สำหรับการวัดการทำหน้าที่ต่างกันของข้อสอบแบบไม่คิดเครื่องหมายในแบบทดสอบที่มีการตรวจให้คะแนนแบบผสมผสาน พบว่า ตัวประมาณค่าการทำหน้าที่ต่างกันของข้อสอบทั่วไปทั้งประเภทถ่วงน้ำหนักและไม่ถ่วงน้ำหนักให้ความแม่นยำในการตรวจสอบและสามารถใช้ในการตรวจสอบการทำหน้าที่ต่างกันของแบบทดสอบของข้อสอบที่มีการตรวจให้คะแนนแบบผสมผสาน (mixed format tests) ได้

อย่างไรก็ตามข้อสอบและแบบสอบที่ดีสามารถสะท้อนคุณภาพการวัดและประเมินผลสัมฤทธิ์ทางการศึกษาที่มีคุณภาพ คุณภาพของข้อสอบและแบบสอบยังสามารถนำมาใช้ในการประเมิน

คุณภาพการจัดการศึกษาที่มีคุณภาพได้เช่นเดียวกัน การตรวจสอบดัชนีคุณภาพของเครื่องมืออย่างต่อเนื่องและจริงจัง ด้วยวิธีการที่หลากหลายสามารถสร้างความน่าเชื่อถือและสร้างการยอมรับในระบบการทดสอบระดับชาติให้เกิดความมั่นใจในการนำผลการประเมินไปใช้ (สุพักตร์ พิบูลย์, 2557) คะแนนจากแบบสอบในหน่วยงานระดับชาติ (national organizations) ที่เกี่ยวข้องกับการวัดและประเมินผลในการจัดการศึกษาของประเทศไทย จากสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) หรือ สทศ. และ สำนักงานรับรองมาตรฐานและประเมินคุณภาพการศึกษา (องค์การมหาชน) หรือ สมศ. ล้วนเป็นหน่วยงานหลักที่ทำหน้าที่ในการทดสอบและประเมินผลการจัดการศึกษาของสถานศึกษาตามมาตรฐานการศึกษาของชาติ เพื่อนำผลการประเมินมาปรับปรุงแก้ไขและพัฒนาคุณภาพการจัดการศึกษาให้กับผู้เรียนและสถานศึกษาในประเทศ นอกจากนี้ยังมีหน่วยงานที่เกี่ยวข้องกับการวัดและประเมินผลการจัดการศึกษาในต่างประเทศ (international organizations) ซึ่งประเทศไทยสามารถนำผลการประเมินดังกล่าวเข้ามาใช้ปรับปรุงกระบวนการจัดการศึกษา เช่น โครงการศึกษาคณิตศาสตร์และวิทยาศาสตร์ระดับนานาชาติ (the international mathematics and science study: TIMSS) เพื่อประเมินผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์และวิทยาศาสตร์ตามหลักสูตรของนักเรียนระดับชั้นประถมศึกษาปีที่ 4 (Grade 4) และมัธยมศึกษาปีที่ 2 (Grade 8) และโครงการประเมินผลนักเรียนนานาชาติ (programme for international student assessment: PISA) เพื่อประเมินคุณภาพของระบบการศึกษาในการเตรียมความพร้อมให้ประชาชนมีศักยภาพหรือความสามารถพื้นฐานที่จำเป็นต่อการดำรงชีวิตในโลกที่มีการเปลี่ยนแปลง ซึ่งโครงการจะเน้นการประเมินสมรรถนะของนักเรียนเกี่ยวกับการใช้ความรู้และทักษะในชีวิตจริงมากกว่าการเรียนรู้ตามหลักสูตรในโรงเรียน

โครงการประเมินผลนักเรียนนานาชาติ หรือ PISA นับเป็นความร่วมมือและความเชี่ยวชาญของนานาชาติเพื่อสร้างชุดของข้อมูลเดียวกันที่มีศักยภาพและชี้บอกคุณภาพของผลการศึกษาของประเทศที่เข้าร่วมโครงการ โดยใช้ทดสอบกับกลุ่มตัวอย่างนักเรียนช่วงอายุ 15 ปี ซึ่งเป็นวัยจบการศึกษาจากทุกสังกัดตามภูมิศาสตร์ของประเทศ รอบการประเมินผลจะดำเนินการสอบทุกๆ สามปี เพื่อติดตามแนวโน้มการเปลี่ยนแปลงคุณภาพการเรียนรู้ของนักเรียนซึ่งจะเป็นข้อมูลสำคัญในระดับนโยบาย โดยเนื้อหาการทดสอบออกเป็น 3 ส่วน ได้แก่ การประเมินผลการรู้เรื่องการอ่าน (reading literacy) การประเมินผลการรู้เรื่องคณิตศาสตร์ (mathematics literacy) และการประเมินรู้เรื่องวิทยาศาสตร์ (scientific literacy) ในปี ค.ศ. 2009 ข้อสอบเน้นไปในเรื่องการอ่านมีน้ำหนัก 60% คณิตศาสตร์และวิทยาศาสตร์มีน้ำหนักอย่างละประมาณ 20% โดยมีจำนวนคำถามประกอบด้วยการอ่าน 103 ข้อ คณิตศาสตร์ 36 ข้อ และวิทยาศาสตร์ 53 ข้อ รวมข้อคำถามทั้งหมด 192 ข้อ

จากผลการประเมินนักเรียนไทยที่เข้าร่วมทดสอบในโครงการประเมินผลนักเรียนนานาชาติ ปี 2009 (สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี, 2553) พบว่า จากประเทศที่เข้าร่วมทั้งหมด 65 ประเทศ ผลการประเมินการรู้เรื่องการอ่านของนักเรียนไทยต่ำกว่าค่าเฉลี่ย OECD โดยมีคะแนนเฉลี่ย 421 อยู่ในช่วงอันดับประมาณ 50 ผลการประเมินการรู้เรื่องคณิตศาสตร์ต่ำกว่าค่าเฉลี่ย OECD โดยมีคะแนนเฉลี่ย 419 อยู่ที่ตำแหน่งอันดับประมาณ 52 และผลการประเมินการรู้เรื่องวิทยาศาสตร์ต่ำกว่าค่าเฉลี่ย OECD โดยมีคะแนนเฉลี่ย 425 อยู่ที่ตำแหน่งอันดับประมาณ 49 เมื่อดูแนวโน้มตั้งแต่ PISA 2000 เป็นต้นมา พบว่า ผลการประเมินมีแนวโน้มต่ำลง สอดคล้องกับผลการประเมินระดับชาติของประเทศไทยจากสำนักทดสอบแห่งชาติ (องค์การมหาชน) หรือ สทศ. จากการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O- NET) ในปีการศึกษา 2555 พบว่า นักเรียนส่วนใหญ่ได้คะแนนต่ำกว่าร้อยละ 50 ในทุกรายวิชายกเว้นวิชาสุขศึกษาและพลศึกษา

คะแนนที่ได้จากแบบทดสอบระดับชาติหรือระดับนานาชาติส่วนใหญ่มักจะถูกนำมาใช้ในการบ่งบอกคุณภาพการจัดการศึกษาของสถานศึกษาในแต่ละประเทศ ดังเช่น สถาบันทดสอบทางการศึกษาแห่งชาติได้นำเสนอผลการวิเคราะห์ค่าความยาก อำนาจจำแนก และค่าความเที่ยงของแบบสอบที่ได้มีการดำเนินการทดสอบซึ่งเกี่ยวข้องกับพันธกิจด้านการประเมินผลการจัดการศึกษาและการทดสอบทางการศึกษาระดับชาติ คะแนนที่ได้จากแบบสอบส่วนใหญ่มักถูกมองข้ามความสำคัญจากประเด็น คะแนนที่ได้จากแบบสอบมักจะเป็นค่าที่ได้จากผลรวมของคะแนนที่แท้จริงกับค่าความคลาดเคลื่อนจากการวัด การนำคะแนนที่ได้จากการสอบมาใช้ตัดสินคุณภาพที่แท้จริงของนักเรียนหรือสถานศึกษาไม่สามารถบอกความสามารถที่แท้จริงของนักเรียนและคุณภาพผลลัพธ์ของกระบวนการศึกษาได้ทั้งหมด เพราะอาจจะมีปัจจัยหลายอย่างซึ่งสะท้อนคุณลักษณะพื้นฐานของนักเรียนและการบริหารจัดการภายในสถานศึกษาที่ส่งผลกระทบต่อการทำคะแนนในแบบสอบที่อาจเกิดขึ้น ปัจจัยที่ส่งผลต่อคุณภาพการจัดการศึกษามักเกิดมาจากแหล่งความคลาดเคลื่อนหลายแหล่ง เช่น ความคลาดเคลื่อนจากการเลือกโรงเรียนเพื่อศึกษาต่อ (selection bias) ความคลาดเคลื่อนจากภายในโรงเรียน (school bias) และความคลาดเคลื่อนจากภูมิศาสตร์ (geographic bias) เป็นต้น การควบคุมตัวแปรหรือปัจจัยในระดับนักเรียนและระดับสถานศึกษาที่เป็นแหล่งความคลาดเคลื่อนจึงส่งผลต่อการเรียนรู้ของนักเรียนที่เป็นความสามารถที่แท้จริง การควบคุมแหล่งความคลาดเคลื่อนดังกล่าวสามารถนำมาใช้เป็นส่วนหนึ่งของกระบวนการประเมินเพื่อทำให้เกิดความเท่าเทียมกันและส่งเสริมให้นักเรียนเกิดการเรียนรู้ที่เพิ่มขึ้น ในปัจจุบันกระบวนการศึกษาดังกล่าวเรียกว่า การวิเคราะห์มูลค่าเพิ่ม (value-added analysis)

การวิเคราะห์มูลค่าเพิ่มหรือโมเดลมูลค่าเพิ่มเป็นวิธีการเปรียบเทียบความสามารถของโรงเรียนและครูในปัจจุบันที่มีความยุติธรรมที่สุด แนวคิด “มูลค่าเพิ่ม” เป็นแนวคิดทางด้าน

เศรษฐศาสตร์ถูกใช้เป็นเครื่องมือวัดผลการดำเนินงานทางธุรกิจในเชิงเศรษฐศาสตร์ มาจากคำว่า economic value added (EVA) พัฒนาขึ้นในช่วงปลายทศวรรษ 1980 โดยบริษัทที่ปรึกษาอเมริกัน Stern Stewart Consulting Group ทั้งนี้ EVA ถูกนำมาใช้เป็นตัววัดทางด้านการเงินเป็นเกณฑ์ในการวัดผลการดำเนินงานและเป็นเครื่องมือสำคัญในการสื่อสารที่มีประสิทธิภาพสูงสุดขององค์กร ภายหลังนักการศึกษาจึงนำแนวคิดดังกล่าวเข้ามาประยุกต์ใช้ทางการศึกษาเพื่อประเมินความสามารถในการดำเนินงานของโรงเรียน โดยการนำแนวคิดมูลค่าเพิ่มมาใช้ในการศึกษามาจากพระราชบัญญัติการศึกษาของประเทศสหรัฐอเมริกา ในยุคศตวรรษที่ 20 ที่เรียกกันว่า no child left behide act of 2001 (NCLB) โดยมีหัวใจสำคัญคือ ผลที่ได้จากการทดสอบทั้งแบบสอบและวิธีการวัดจะต้องสามารถตรวจสอบได้ (accountability) ถึงแม้ว่าโมเดลมูลค่าเพิ่มจะมีหลากหลายรูปแบบในการตรวจสอบแต่รูปแบบที่ได้รับความนิยมและมีการพัฒนาให้เป็นที่รู้จัก ก็คือการวัดมูลค่าเพิ่มของเทนเนสซี (tennessess value-added assessment system: TVAAS) พัฒนามาจาก William L. Sanders และผู้ช่วย เป็นระบบการประเมินที่ใช้ในรัฐเทนเนสซีของประเทศสหรัฐอเมริกาสำหรับออกแบบเพื่อวัดผลกระทบของสถานศึกษาและครูที่ส่งผลต่อความสำเร็จของนักเรียน ภายหลังได้มีการพัฒนาและอาศัยวิธีทางสถิติเข้ามาตรวจสอบคะแนนผลสัมฤทธิ์ของนักเรียนระดับกลุ่มขนาดใหญ่ (large-scale) ซึ่งต่อมาภายหลังถูกเรียกว่า การประเมินมูลค่าเพิ่มทางการศึกษา (education value-added assessment system: EVAAS) (Battele of Kids, 2007; Sanders, 1998; Olson, 2004 cited in Beardsley, 2014)

การประเมินมูลค่าเพิ่มทางการศึกษา เป็นการศึกษาการเรียนรู้ซึ่งอาศัยการเปรียบเทียบระหว่างคะแนนผลการเรียนรู้ที่เกิดขึ้นจริงหรือสังเกตได้ (observed scores) กับคะแนนผลการเรียนรู้ที่ทำนายได้ (predicted scores) จากตัวแปรภูมิหลังของนักเรียน บริบทชุมชน สังคม หรือผลสัมฤทธิ์เดิม (Manzi et al., 2010; ศิริชัย กาญจนวาสี, 2550ข) นอกจากนี้ประสิทธิภาพของโมเดลมูลค่าเพิ่มยังช่วยขจัดความลำเอียงที่สามารถอธิบายการดำเนินงานของครูและโรงเรียนได้อย่างแม่นยำ อีกทั้งคะแนนมูลค่าเพิ่มยังมีประโยชน์ต่อนักวิจัยสามารถนำมาประเมินมาจัดอันดับ (ranking) ประสิทธิผลของโรงเรียนโดยการควบคุมตัวแปรคุณลักษณะของนักเรียนและคุณลักษณะโรงเรียนหรือสิ่งแวดล้อม McCaffrey และคณะ (2003) ได้กล่าวว่า สำหรับแต่ละโรงเรียนการลดความคลาดเคลื่อนที่เกิดขึ้นควรรวมตัวแปรที่เกี่ยวข้องเป็นตัวแปรพยากรณ์ในโมเดลถดถอยมูลค่าเพิ่ม (value-added regression models) หรือโมเดลอิทธิพลคงที่ของโรงเรียน (school fixed effects model) ซึ่งการรวมหรือแยกตัวแปรมักส่งผลไปยังอิทธิคงที่ของโรงเรียนและกระทบต่อประสิทธิผลของครูตามมา

การนำเอาแนวคิดมูลค่าเพิ่มทางการศึกษามาใช้ในวงการศึกษานับว่าเป็นเรื่องที่ใหม่และกำลังอยู่ในความสนใจทั้งในประเทศและต่างประเทศ การนำแนวคิดการวิเคราะห์มูลค่าเพิ่มมาใช้ใน

การประเมินคุณภาพการจัดการศึกษา จึงเป็นอีกแนวทางที่สามารถให้ข้อสารสนเทศในการปฏิบัติงานแก่สถานศึกษา อีกทั้งยังเป็นประโยชน์ต่อการใช้การกำหนดนโยบายทั้งต่อผู้บริหาร ครู ตลอดจนผู้ปกครองและนักเรียนได้ (OECD, 2008) นอกจากนี้ในงานวิจัยที่เกี่ยวข้องมูลค่าเพิ่มทางการศึกษาส่วนใหญ่มักเป็นงานวิจัยที่เกี่ยวข้องกับการวัดผลสัมฤทธิ์ของนักเรียนในสถานศึกษา ในต่างประเทศ เช่น Ready (2013) ศึกษาความสัมพันธ์ระหว่างผลสัมฤทธิ์ของนักเรียนและการเรียนรู้ Keeves, Hungi, และ Afrassa (2005) ศึกษาการวัดผลมูลค่าเพิ่มข้ามโรงเรียน ในแบบทดสอบทักษะพื้นฐานการอ่านและตัวเลข ส่วน Noell และ Burns (2008) วัดมูลค่าเพิ่มของการเตรียมการเป็นครูในรัฐลุสเซียน่า ระหว่างปี 2004-2005 ถึง 2006-2007 โดยใช้การวิเคราะห์ลดหลั่นแบบซ้อนทับ (nested) ซึ่งนักเรียนอยู่ภายในอาจารย์และอาจารย์อยู่ภายในโรงเรียน ส่วนในประเทศไทย สุชีรา มะหิเมือง (2547) ศึกษาปัจจัยที่ส่งผลต่อผลสัมฤทธิ์และพัฒนาการทางวิชาการ โดยการวิเคราะห์มูลค่าเพิ่ม และจตุภูมิ เขตจัตุรัส (2552) พัฒนาโมเดลมูลค่าเพิ่มของผลสัมฤทธิ์ทางวิชาการและแบบตรวจสอบรายการประเมินตนเองเพื่อเพิ่มมูลค่ากระบวนการจัดการศึกษา ส่วนเพ็ญภัทร พื้นผา (2554) ได้วิเคราะห์โมเดลมูลค่าเพิ่มพหุระดับเพื่อการจัดประสิทธิผลของโรงเรียน

นอกจากงานวิจัยมูลค่าเพิ่มทางการศึกษาที่เกี่ยวข้องกับการวัดผลสัมฤทธิ์ของนักเรียนภายในสถานศึกษา ยังมีงานวิจัยส่วนหนึ่งที่เล็งเห็นประโยชน์ในการนำการวิเคราะห์มูลค่าเพิ่มเข้ามาศึกษาในบริบททางด้านการวัดผลและประเมินผลระดับการประเมินขนาดกลุ่มใหญ่ (large scale assessment) โดยพัฒนาจากโมเดลการวิเคราะห์มูลค่าเพิ่มพหุระดับ ซึ่งงานวิจัยดังกล่าวให้ความสำคัญกับการศึกษาความยุติธรรมของข้อสอบโดยการขจัดความลำเอียงของข้อสอบบางข้อออกไป แนวคิดดังกล่าวเป็นแนวคิดที่สร้างคุณค่าให้กับศาสตร์ทางด้านการวัดผลและประเมินผลเป็นอย่างมาก อีกทั้งการนำแนวคิดมูลค่าเพิ่มเข้ามาประยุกต์ใช้ทำให้เกิดองค์ความรู้ทางด้านวัดผลและประเมินผลเพิ่มมากขึ้นในปัจจุบัน งานวิจัยที่เกี่ยวข้องกับข้อสอบและการทำหน้าที่ต่างกันของข้อสอบ เช่น งานวิจัยของ Tuksino, Kanjanawasee, และ Pasiphol (2010) ได้ประยุกต์ใช้โมเดลมูลค่าเพิ่ม (value-added model) ด้วยโมเดลเชิงเส้นตรงแบบลดหลั่น (HLM) ที่ได้มาจากการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบเพื่อประเมินคุณภาพการจัดการศึกษาวิชาวิทยาศาสตร์ โดยการตรวจสอบผลการประเมินคุณภาพการจัดการศึกษาของสถานศึกษาจากโมเดลที่ต่างกัน 4 โมเดล พบว่า โมเดลประเมินคุณภาพการจัดการศึกษาวิชาวิทยาศาสตร์ที่ตัดข้อสอบทำหน้าที่ต่างกันออกและวิเคราะห์มูลค่าเพิ่มโดยการปรับแก้คะแนนด้วยตัวแปรระดับนักเรียนและสถานศึกษาที่มีประสิทธิ์ในการทำนายสูงสุด เช่นเดียวกับ Sirikit, Pasiphol, และ Kanjanawasee (2012) ได้ประยุกต์ใช้โมเดลมูลค่าเพิ่มที่ได้มาจากการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและการวิเคราะห์การทำหน้าที่ต่างกันของตัวลวงเพื่อการประเมินคุณภาพการจัดการศึกษาวิชาคณิตศาสตร์ โดยการตรวจสอบผลการประเมินคุณภาพการจัดการศึกษาของสถานศึกษาจากโมเดลที่ต่างกัน 4 โมเดล

พบว่า โมเดลประเมินคุณภาพการจัดการศึกษาคณิตศาสตร์ที่ตัดข้อสอบที่ทำหน้าที่ต่างกันโดยการวิเคราะห์ทั้งตัวถูกและตัวลวงและวิเคราะห์มูลค่าเพิ่มโดยการควบคุมตัวแปรระดับนักเรียนและระดับสถานศึกษามีประสิทธิ์ในการทำนายสูงสุด ส่วนงานวิจัยที่เกี่ยวข้องกับแบบสอบ เช่นงานวิจัยของ Papay (2011) ได้วัดผลลัพธ์การประมาณค่ามูลค่าเพิ่มของครูโดยพิจารณาแบบสอบที่ต่างกันและคำตอบในการตอบที่แตกต่างกันในแบบทดสอบวัดผลสัมฤทธิ์ทางการอ่าน 3 ฉบับ ได้แก่ แบบทดสอบระดับรัฐ (state test) แบบทดสอบวัดผลสัมฤทธิ์ของสแตนฟอร์ด (standford achievement test: SAT) และเครื่องมือวัดการอ่านในโรงเรียน (scholastic reading inventory: SRT) พบว่า แบบทดสอบทั้ง 3 ฉบับมีสหสัมพันธ์แบบเรียงอันดับอยู่ในระดับขนาดกลาง (moderate-sized rank) ซึ่งมีค่าตั้งแต่ 0.15 – 0.58 แต่โมเดลที่ในการศึกษาโมเดลมูลค่าเพิ่ม 9 โมเดลที่แตกต่างกันสะท้อนผลการประเมินความสามารถของครูที่แตกต่างกัน

ดังนั้นจากเอกสารและงานวิจัยที่เกี่ยวข้องจะเห็นได้ว่างานวิจัยที่เกี่ยวข้องกับการวิเคราะห์มูลค่าเพิ่มทางการศึกษาทางด้านการวัดผลและประเมินผลมีค่อนข้างน้อย เมื่อเปรียบเทียบกับการวัดผลสัมฤทธิ์ทางการเรียนภายในสถานศึกษา จากเหตุผลดังกล่าว ผู้วิจัยจึงมีความสนใจที่จะศึกษาเปรียบเทียบโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์โดยประยุกต์ใช้โมเดลมูลค่าเพิ่มพหุระดับที่มีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ โดยใช้ข้อมูลจากโครงการประเมินผลนักเรียนนานาชาติ (PISA) ค.ศ. 2009 ซึ่งเป็นแบบทดสอบที่มีข้อสอบทั้งประเภทตรวจให้คะแนนทวิวิภาคและพหุวิภาค และเป็นโครงการที่ใช้ผลคะแนนที่ได้จากการทำแบบทดสอบโดยไม่คำนึงความแตกต่างของคุณลักษณะพื้นฐานในระดับนักเรียนและระดับสถานศึกษาในประเทศไทย นอกจากนี้ผู้วิจัยยังขยายแนวคิดการตัดข้อสอบที่มีการตรวจให้คะแนนพหุวิภาคเข้ามาร่วมวิเคราะห์โดยใช้การศึกษาการทำหน้าที่ต่างกันเป็นขั้น (DSF) และเพิ่มการตรวจสอบการทำหน้าที่ต่างกันของแบบทดสอบ (DTF) โดยพิจารณาถึงโครงสร้างของแบบทดสอบในโครงการประเมินผลนักเรียนนานาชาติที่มีการตรวจให้คะแนนแบบผสมผสาน (mixed format tests) นอกจากนี้ผู้วิจัยยังใช้เทคนิควิธีการวิเคราะห์ข้อมูลโดยประยุกต์ใช้การวิเคราะห์มูลค่าเพิ่มพหุระดับ 2 ระดับ คือ ระดับนักเรียนและระดับสถานศึกษา และเพื่อเพิ่มความน่าเชื่อถือและความยุติธรรมของผลการวัดโดยการเปรียบเทียบโมเดลการวัด 2 โมเดล คือ โมเดลที่ 1 เป็นโมเดลการประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ที่ไม่ได้พิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ ส่วนโมเดลการวิเคราะห์มูลค่าเพิ่มพหุระดับเป็นโมเดลที่มีการควบคุมตัวแปรระดับนักเรียนและสถานศึกษา และโมเดลที่ 2 เป็นโมเดลการประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ที่มีการพิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบโดยการตัดข้อสอบที่มีการทำหน้าที่ต่างกันของข้อสอบและการทำหน้าที่ต่างกันเป็นขั้นที่แตกต่างกัน ส่วนโมเดลการวิเคราะห์มูลค่าเพิ่มพหุระดับเป็นโมเดลที่มีควบคุมตัวแปรระดับนักเรียนและสถานศึกษา ซึ่งเป็นการพัฒนาองค์ความรู้และนำเสนอสารสนเทศที่

เป็นประโยชน์ทางด้านวัดผลการศึกษา อีกทั้งยังเป็นการขยายองค์ความรู้ทางด้านการวิเคราะห์มูลค่าเพิ่มอันจะเป็นประโยชน์กับผู้เกี่ยวข้องทุกฝ่ายที่จะได้รับความรู้ในด้านวิชาการและการนำความรู้ไปใช้ปฏิบัติจริง

คำถามการวิจัย

1. ผลการทำหน้าที่ต่างกันของข้อสอบ(DIF) ที่มีการตรวจให้คะแนนทวิวิภาคและพหุวิภาค และการทำหน้าที่ต่างกันของแบบทดสอบ (DTF) ในแบบทดสอบการประเมินการรู้เรื่องวิทยาศาสตร์มากน้อยเพียงใด และเป็นอย่างไร

2. ประสิทธิภาพของโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์เมื่อใช้โมเดลมูลค่าเพิ่มพหุระดับ 2 โมเดล คือ โมเดลที่ไม่ได้พิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ (โมเดลที่ 1) กับโมเดลที่มีการพิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบโดยการตัดข้อสอบที่มีการทำหน้าที่ต่างกันของข้อสอบและการทำหน้าที่ต่างกันเป็นขั้นที่แตกต่างกัน (โมเดลที่ 2) ที่มีผลต่อคุณภาพการจัดการศึกษาแตกต่างกันหรือไม่ เพียงใด

วัตถุประสงค์การวิจัย

1. เพื่อศึกษาผลการทำหน้าที่ต่างกันของข้อสอบ (DIF) ที่มีการตรวจให้คะแนนทวิวิภาคและพหุวิภาค และการทำหน้าที่ต่างกันของแบบทดสอบ (DTF) ในแบบทดสอบการประเมินการรู้เรื่องวิทยาศาสตร์

2. เพื่อเปรียบเทียบประสิทธิภาพของโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์เมื่อใช้โมเดลมูลค่าเพิ่มพหุระดับ 2 โมเดล คือ โมเดลที่ไม่ได้พิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ (โมเดลที่ 1) กับโมเดลที่มีการพิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบโดยการตัดข้อสอบที่มีการทำหน้าที่ต่างกันของข้อสอบและการทำหน้าที่ต่างกันเป็นขั้นที่แตกต่างกัน (โมเดลที่ 2) ซึ่งในการเปรียบเทียบประสิทธิภาพของโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ มีวัตถุประสงค์ย่อยคือ

2.1 เพื่อศึกษาความสัมพันธ์ระหว่างคะแนนมูลค่าเพิ่มกับ 3 ตัวแปร คือ เศรษฐานะของครอบครัว แหล่งทรัพยากรการศึกษาที่บ้าน และความมั่งคั่งของครอบครัว ทั้งก่อนและหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน

2.2 เพื่อเปรียบเทียบความสอดคล้องของการจัดอันดับคุณภาพและการจัดกลุ่มคุณภาพของโมเดลมูลค่าเพิ่มพหุระดับทั้ง 2 โมเดล

สมมติฐานการวิจัย

ข้อสอบที่ตรวจพบทำหน้าที่ต่างกันอาจมีแหล่งการทำหน้าที่ต่างกันหลากหลายแหล่ง เนื่องมาจากกลุ่มผู้เข้าสอบและระดับข้อมูลที่แตกต่างทำให้เกิดความไม่เป็นธรรมกับผู้เข้าสอบบางกลุ่ม ดังนั้นข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันจะมีแหล่งของตัวแปรในแต่ละระดับที่นำมาวิเคราะห์ซึ่งส่งผลต่อความไม่เป็นธรรมกับผู้เข้าสอบบางกลุ่มได้ นอกจากนี้แบบทดสอบในปัจจุบันมีทั้งแบบทดสอบที่ประกอบไปด้วยข้อสอบที่มีการตรวจให้คะแนนทวิวิภาคและพหุวิภาค หรือข้อสอบที่มีการตรวจให้คะแนนแบบผสมผสานเพื่อใช้ในการวัดและประเมินผลการเรียนรู้ นอกจากนี้จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการประเมินคุณภาพการจัดการศึกษา พบว่า ความไม่เท่าเทียมกันในสถานศึกษาอาจจะเกิดจากแหล่งความคลาดเคลื่อนที่มาจากต่างแหล่ง เช่น ความคลาดเคลื่อนจากการเลือกโรงเรียนเพื่อศึกษาต่อ (selection bias) จากผู้ปกครอง ความคลาดเคลื่อนจากภายในโรงเรียน (school bias) และความคลาดเคลื่อนจากภูมิศาสตร์ (geographic bias) เมื่อโรงเรียนที่แตกต่างกันอาจจะมีแหล่งความความเคลื่อนที่ต่างกัน นอกจากนี้การวิเคราะห์พหุระดับสามารถควบคุมตัวแปรหรือปัจจัยในระดับนักเรียนและระดับสถานศึกษาที่เป็นแหล่งความคลาดเคลื่อนซึ่งมีผลต่อการเรียนรู้ของนักเรียน การวิเคราะห์พหุระดับที่มีการศึกษา คือ การวิเคราะห์มูลค่าเพิ่ม (value-added analysis) เพื่อลดความคลาดเคลื่อนและทำให้เกิดความเท่าเทียมกัน ซึ่งจะส่งเสริมให้เกิดความก้าวหน้าของนักเรียนและเป็นไปตามวัตถุประสงค์การจัดการศึกษาได้

จากงานวิจัยของ Ready (2013), Keeves, Hungi, และ Afrassa (2005), Noell และ Burns (2008), สุชีรา มะหิเมือง (2547), และจตุภูมิ เขตจัตุรัส (2552) เน้นไปที่การศึกษาผลสัมฤทธิ์ของนักเรียนมีผลต่อการประเมินคุณภาพโรงเรียน ส่วน Tuksino, Kanjanawasee, และ Pasiphol (2010) Sirikit, Pasiphol, และ Kanjanawasee (2012) และ Papay (2011) เน้นไปที่การศึกษามูลค่าเพิ่มทางด้านการวัดผลและประเมินผลในข้อสอบ อีกทั้ง Tuksino, Kanjanawasee, และ Pasiphol (2010) และ Sirikit, Pasiphol, และ Kanjanawasee (2012) ยังได้ศึกษาเปรียบเทียบความสอดคล้องระหว่างโมเดลข้อสอบก่อนกับหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่าโมเดลข้อสอบก่อนตัดข้อสอบกับหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันสอดคล้องกัน ดังนั้นจากเหตุผลและงานวิจัยที่เกี่ยวข้องโมเดลคุณภาพการจัดการศึกษาวิทยาศาสตร์ที่มีประสิทธิภาพและมีความยุติธรรมที่สุดน่าจะเป็นการประยุกต์ใช้โมเดลวิเคราะห์มูลค่าเพิ่มที่ควบคุมอิทธิพลของตัวแปรในระดับนักเรียนและระดับสถานศึกษาเพื่อความเท่าเทียมกันที่อาจจะส่งผลต่อคุณภาพการจัดการศึกษาพร้อมกับการพิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ ทำให้ผู้วิจัยจึงกำหนดสมมติฐานการวิจัย 2 ข้อดังนี้

1. ผลการทำหน้าที่ต่างกันของข้อสอบ (DIF) ที่มีการตรวจให้คะแนนทวิวิภาคและพหุวิภาค และการทำหน้าที่ต่างกันของแบบทดสอบ (DTF) ในแบบทดสอบการประเมินการรู้เรื่องวิทยาศาสตร์ ทั้งสองโมเดลมีความแตกต่างกัน

2. ประสิทธิภาพของโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์เมื่อใช้โมเดลมูลค่าเพิ่มพหุระดับ 2 โมเดล คือ โมเดลที่ไม่ได้พิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ (โมเดลที่ 1) กับโมเดลที่มีการพิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบโดยการตัดข้อสอบที่มีการทำหน้าที่ต่างกันของข้อสอบและการทำหน้าที่ต่างกันเป็นขั้นที่แตกต่างกัน (โมเดลที่ 2) ซึ่งโมเดลที่มีประสิทธิภาพที่สุด คือ โมเดลที่มีการพิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบโดยการตัดข้อสอบที่มีการทำหน้าที่ต่างกันของข้อสอบและการทำหน้าที่ต่างกันเป็นขั้นที่แตกต่างกัน (โมเดลที่ 2) โดยที่

2.1 คะแนนมูลค่าเพิ่มกับ 3 ตัวแปร คือ เศรษฐานะของครอบครัว แหล่งทรัพยากรที่บ้าน และความมั่งคั่งของครอบครัว ทั้งก่อนและหลังตัดข้อสอบที่ทำหน้าที่ต่างกันมีความสัมพันธ์กันและคะแนนมูลค่าเพิ่มกับทั้ง 3 ตัวแปรภายหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันออกไปแล้วในการวิเคราะห์มูลค่าเพิ่มมีขนาดลดลงเมื่อเปรียบเทียบกับการวิเคราะห์ที่ไม่ได้ตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันออกไป

2.2 ความสอดคล้องของการจัดอันดับคุณภาพและการจัดกลุ่มคุณภาพของโมเดลมูลค่าเพิ่มพหุระดับทั้ง 2 โมเดลมีความสอดคล้องกัน

ขอบเขตการวิจัย

1) ข้อมูลที่ใช้ในการวิจัยครั้งนี้เป็นข้อมูลทุติภูมิจากฐานข้อมูล (data base) ของโครงการประเมินผลนักเรียนนานาชาติ หรือ PISA (Programme for International Student Assessment) ปี ค.ศ. 2009 กลุ่มตัวอย่างของโครงการเป็นนักเรียนที่มีอายุ 15 ในทางปฏิบัติได้ใช้นักเรียนที่มีช่วงอายุ 15 ปี 2 เดือน จนถึง 16 ปี 3 เดือน จำนวน 6,612 คน จากโรงเรียนกลุ่มตัวอย่าง 230 โรงเรียนที่สุ่มไว้ ซึ่งครอบคลุมสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน สำนักงานคณะกรรมการการศึกษาเอกชน สำนักงานการศึกษากรุงเทพมหานคร สำนักประสานและพัฒนา การจัดการศึกษาท้องถิ่น สำนักงานคณะกรรมการการอุดมศึกษา และสำนักงานคณะกรรมการ การอาชีวศึกษา

2) แบบทดสอบที่ใช้ในการศึกษา เป็นแบบทดสอบที่เน้นการประเมินการรู้เรื่องวิทยาศาสตร์ (science Literacy) ประกอบด้วย 4 องค์ประกอบที่มีความสัมพันธ์ซึ่งกันและกัน ได้แก่ บริบทของการประเมินผลวิทยาศาสตร์ ได้แก่ การรับรู้ถึงสถานการณ์ในชีวิตที่เกี่ยวข้องกับวิทยาศาสตร์และเทคโนโลยี สมรรถนะทางวิทยาศาสตร์ หมายถึง การใช้ความรู้วิทยาศาสตร์ใน

การระบุประเด็นวิทยาศาสตร์ (identifying scientific issues) การอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ (explain phenomena scientifically) และการใช้ประจักษ์พยานทางวิทยาศาสตร์ (using scientific evidence) ความรู้ทางวิทยาศาสตร์ประกอบด้วยสองส่วน คือ ความรู้ในเรื่องโลกธรรมชาติเรียกว่า ความรู้วิทยาศาสตร์ กับความรู้ที่เกี่ยวข้องกับความรู้วิทยาศาสตร์นั้นๆ เรียกว่า ความรู้เกี่ยวกับวิทยาศาสตร์ และเจตคติเชิงวิทยาศาสตร์ ได้แก่ การแสดงการตอบสนองต่อวิทยาศาสตร์ด้วยความสนใจสนับสนุนการสืบหาความรู้วิทยาศาสตร์ และแสดงความรับผิดชอบต่อสิ่งต่างๆ เช่น ในประเด็นของทรัพยากรธรรมชาติและสิ่งแวดล้อม เป็นต้น ประกอบไปด้วยข้อสอบทั้งหมด 53 ข้อ แบบเลือกตอบธรรมดา 18 ข้อ เลือกตอบเชิงซ้อน 17 ข้อ สร้างคำตอบอิสระ 17 ข้อ และสร้างคำตอบปิด 1 ข้อ ทั้งนี้แบบทดสอบมีทั้งหมด 13 booklets ส่วนแบบทดสอบที่ใช้ประเมินการรู้เรื่องวิทยาศาสตร์มีจำนวน 9 booklets โดยที่นักเรียนหนึ่งคนไม่ได้ตอบทั้งหมด 53 ข้อ แต่ละถูกสุ่มให้ได้รับแบบทดสอบฉบับใดฉบับหนึ่งซึ่งมีความยากง่ายพอๆ กัน ซึ่งจะถูกประเมินความสามารถด้านการอ่าน คณิตศาสตร์ และวิทยาศาสตร์ ซึ่งในแต่ละ booklet มีข้อสอบซ้ำกันอยู่บ้าง ไม่ซ้ำกันบ้าง แบบทดสอบบางฉบับมีทั้งข้อสอบการอ่าน คณิตศาสตร์ และวิทยาศาสตร์ บางฉบับมีเพียงข้อสอบการอ่านและคณิตศาสตร์ หรือวิทยาศาสตร์ และบางฉบับมีเฉพาะขอสอบการอ่านอย่างเดียว กำหนดเวลาทำข้อสอบ 2 ชั่วโมง

3) โมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์กำหนดเป็นโมเดลในการประเมินเป็น 2 โมเดลดังต่อไปนี้

โมเดลที่ 1 เป็นโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ที่ไม่ได้พิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ ส่วนโมเดลการวิเคราะห์มูลค่าเพิ่มพหุระดับเป็นโมเดลที่มีควบคุมตัวแปรระดับนักเรียนและสถานศึกษา

โมเดลที่ 2 เป็นโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ที่พิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ โดยการตัดข้อสอบที่มีการทำหน้าที่ต่างกันของข้อสอบและการทำหน้าที่ต่างกันเป็นขั้นที่แตกต่างกัน ส่วนโมเดลการวิเคราะห์มูลค่าเพิ่มพหุระดับเป็นโมเดลที่มีการควบคุมตัวแปรระดับนักเรียนและสถานศึกษา

คำจำกัดความที่ใช้ในการวิจัย

แบบทดสอบสำหรับประเมินความรู้เรื่องวิทยาศาสตร์ หมายถึง แบบทดสอบการรู้เรื่องวิทยาศาสตร์ ซึ่งวัดความรู้เรื่องวิทยาศาสตร์ตามกรอบการประเมินความรู้เรื่องวิทยาศาสตร์ในโครงการประเมินผลนักเรียนนานาชาติ (programme for international student assessment: PISA) ปี ค.ศ. 2009 จำนวน 9 ฉบับ โดยแบ่งเป็นความรู้วิทยาศาสตร์ ได้แก่ ระบบโลกและอวกาศ ระบบการดำรงชีวิต ระบบทางกายภาพ ระบบเทคโนโลยี และวิทยาศาสตร์และเทคโนโลยีในสังคม และความรู้เกี่ยวกับวิทยาศาสตร์ ได้แก่ การอธิบายเชิงวิทยาศาสตร์ และการเสาะหาความรู้เชิงวิทยาศาสตร์

คุณภาพการจัดการศึกษาวิทยาศาสตร์ หมายถึง คะแนนมูลค่าเพิ่มของสถานศึกษาหรือค่าส่วนที่เหลือจากการวิเคราะห์มูลค่าเพิ่มพหุระดับในแต่ละโมเดลการประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ ซึ่งคุณภาพการจัดการศึกษาของสถานศึกษาจากการวิจัยครั้งนี้ แสดงถึงการดำเนินงานจัดการของแต่ละสถานศึกษาเพื่อให้นักเรียนเกิดการเรียนรู้ที่มีการควบคุมปัจจัยอื่นที่อยู่นอกเหนือการควบคุมของนักเรียนและสถานศึกษาออกไปแล้ว

การประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ หมายถึง การตัดสินคุณค่าของสถานศึกษาขั้นพื้นฐานโดยการวิเคราะห์หาคุณภาพการจัดการศึกษาของสถานศึกษา จากการใช้ผลการทำแบบทดสอบสำหรับประเมินการรู้เรื่องวิทยาวิทยาศาสตร์จากโครงการประเมินผลนักเรียนนานาชาติ

โมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ หมายถึง โมเดลที่ใช้ในการประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์โดยการประยุกต์ใช้โมเดลมูลค่าเพิ่มพหุระดับที่มีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ ประกอบด้วย 2 โมเดล ได้แก่ **โมเดลที่ 1** เป็นโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ที่ไม่ได้พิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ ส่วนโมเดลการวิเคราะห์มูลค่าเพิ่มพหุระดับเป็นโมเดลมีการควบคุมตัวแปรระดับนักเรียนและสถานศึกษา และ **โมเดลที่ 2** เป็นโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ที่พิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ โดยการตัดข้อสอบที่มีการทำหน้าที่ต่างกันของข้อสอบและการทำหน้าที่ต่างกันเป็นขั้นที่แตกต่างกัน ส่วนโมเดลการวิเคราะห์มูลค่าเพิ่มเป็นโมเดลที่มีควบคุมตัวแปรระดับนักเรียนและสถานศึกษา

คะแนนความสามารถของนักเรียนจากการประเมินผลการรู้เรื่องวิทยาศาสตร์ หมายถึง ค่าพารามิเตอร์จากการวิเคราะห์โมเดลโลจิสติก 3 พารามิเตอร์ ได้มาจากการประมาณค่าความสามารถของนักเรียนจากการทดสอบในโครงการประเมินผลนักเรียนนานาชาติ (PISA) ปี ค.ศ. 2009 โดยโปรแกรม IRTPRO Version 2.1

การเปรียบเทียบความสอดคล้องโมเดลการประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ หมายถึง การเปรียบเทียบความสอดคล้องการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่ม (VAS ranking) โดยที่การจัดอันดับคุณภาพคะแนนมูลค่าเพิ่มใช้การเรียงลำดับคะแนนมูลค่าเพิ่มจากมากที่สุดไปน้อยที่สุด และการเปรียบเทียบความสอดคล้องการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม (VAS rating) การรู้เรื่องวิทยาศาสตร์ โดยการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่มแบ่งเป็น 5 ระดับคุณภาพ ได้แก่ ระดับ A (ดีมาก) ระดับ B (ดี) ระดับ C (พอใช้) ระดับ D (ผ่าน) และระดับ E (ควรปรับปรุง)

เพศ หมายถึง ลักษณะบุคลิกภาพของนักเรียนที่แสดงถึงความเป็นชายหรือความเป็นหญิง ที่ได้มาจากการตอบแบบสอบถามของนักเรียนที่เข้าร่วมการประเมินโครงการประเมินผลนักเรียนนานาชาติ

การเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา หมายถึง การเรียนรู้เนื้อหาหรือกิจกรรมภายนอกห้องเรียนหรือสถานศึกษาตามความสนใจเฉพาะผู้เรียนในวิชาวิทยาศาสตร์ ซึ่งมาจากแบบสอบถามของนักเรียนที่เข้าร่วมการประเมินโครงการประเมินผลนักเรียนนานาชาติ

เศรษฐานะของครอบครัว หมายถึง ดัชนีชี้วัดที่ได้มาจากจากคะแนนรวมที่วัดได้จากสถานะทางการงาน การศึกษาของพ่อแม่ ฐานะทางเศรษฐกิจ ทรัพยากรด้านการศึกษาและวัฒนธรรมของครอบครัวของเด็กนักเรียน ซึ่งมาจากแบบสอบถามของนักเรียนที่เข้าร่วมการประเมินโครงการประเมินผลนักเรียนนานาชาติ

แหล่งทรัพยากรการศึกษาที่บ้าน หมายถึง ดัชนีชี้วัดที่ได้มาจากคะแนนรวมที่วัดได้จากรายการสิ่งของที่มีอยู่ภายในบ้าน เช่น พจนานุกรม คอมพิวเตอร์ โต๊ะทำงาน หนังสือเรียน ซึ่งมาจากแบบสอบถามของนักเรียนที่เข้าร่วมการประเมินโครงการประเมินผลนักเรียนนานาชาติ

ความมั่งคั่งของครอบครัว หมายถึง ดัชนีชี้วัดที่ได้มาจากคะแนนรวมที่วัดได้จากสิ่งของเครื่องใช้ต่างๆ ภายในบ้าน เช่น อินเทอร์เน็ต เครื่องล้างจาน เครื่องเล่นดีวีดี โทรศัพท์เคลื่อนที่ โทรทัศน์ รถ และห้องน้ำในบ้านที่มีฝักบัวหรืออ่างอาบน้ำ ซึ่งมาจากแบบสอบถามของนักเรียนที่เข้าร่วมการประเมินโครงการประเมินผลนักเรียนนานาชาติ

การทำหน้าที่ต่างกันของข้อสอบ (Differential item functioning: DIF) หมายถึง ข้อสอบที่ผู้สอบต่างกลุ่มกันที่มีระดับความสามารถของผู้เข้าสอบเท่ากัน แต่มีโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกัน ในการวิจัยครั้งนี้ใช้การวิเคราะห์ข้อสอบที่ทำหน้าที่ต่างกันโดยตรวจสอบคุณลักษณะของนักเรียนจาก 5 ตัวแปร ได้แก่ เพศ การเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา เศรษฐานะของครอบครัว แหล่งทรัพยากรการศึกษาที่บ้าน และความมั่งคั่งของครอบครัว

การทำหน้าที่ต่างกันของแบบทดสอบ (Differential test functioning: DTF) หมายถึง ลักษณะของแบบสอบที่ทำให้ผลการตอบของผู้สอบที่มีความสามารถเท่ากันในสิ่งที่ต้องการวัดได้คะแนนจริงไม่เท่ากัน เนื่องจากผู้สอบอยู่ในกลุ่มย่อยต่างกัน ในการวิจัยครั้งนี้ใช้การวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบจากคุณลักษณะของนักเรียนจาก 5 ตัวแปร ได้แก่ เพศ การเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา เศรษฐานะของครอบครัว แหล่งทรัพยากรการศึกษาที่บ้าน และความมั่งคั่งของครอบครัว

การทำหน้าที่ต่างกันเป็นขั้น (Differential Step Functioning: DSF) หมายถึง วิธีที่ใช้ในการตรวจสอบความแตกต่างระหว่างกลุ่ม (between-group differences) ในแต่ละขั้น (step) ของข้อสอบที่มีการตรวจให้คะแนนพหุวิภาค (Polytomous items) ทั้งนี้ข้อสอบที่มีการตรวจให้คะแนนพหุวิภาค ที่มี r ลำดับ สามารถกำหนดการทำหน้าที่เป็นขั้นได้ทั้งหมด j = r-1 ขั้น โดยแต่ละขั้นจะแทนด้วยสัญลักษณ์ J ซึ่ง j มีค่าตั้งแต่ j = 1, 2, ... , J ในการวิจัยครั้งนี้ใช้การวิเคราะห์ข้อสอบที่ทำหน้าที่ต่างกันเป็นขั้นโดยการตรวจสอบจากคุณลักษณะของนักเรียน 5 ตัวแปร ได้แก่ เพศ การเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา เศรษฐานะของครอบครัว แหล่งทรัพยากรการศึกษาที่บ้าน และความมั่งคั่งของครอบครัว

การวิเคราะห์พหุระดับ หมายถึง วิธีการตรวจสอบคุณภาพของข้อสอบ โดยลักษณะข้อมูลมีลักษณะที่ข้อสอบสอดแทรกกับผู้สอบแต่ละคน และผู้สอบยังสอดแทรกอยู่ในสถานศึกษา การวิเคราะห์จึงจัดเป็นการวิเคราะห์ 2 ระดับ คือ ระดับที่ 1 คือ ระดับผู้สอบ และระดับที่ 2 คือ ระดับสถานศึกษา

ตัวแปรต้นระดับนักเรียนในการวิเคราะห์พหุระดับ หมายถึง คุณลักษณะของนักเรียนที่นอกเหนือจากความสามารถของผู้เข้าสอบที่นำมาศึกษาในโมเดลมูลค่าเพิ่มพหุระดับ ได้แก่ เพศ เศรษฐานะของครอบครัว แหล่งทรัพยากรการศึกษาที่บ้าน ความมั่งคั่งของครอบครัว แหล่งทรัพยากรทางไอซีที และการใช้เวลาเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา

ตัวแปรต้นระดับสถานศึกษาในการวิเคราะห์พหุระดับ หมายถึง คุณลักษณะของสถานศึกษาขั้นพื้นฐานที่นำมาศึกษาในโมเดลมูลค่าเพิ่มพหุระดับ ได้แก่ ขนาดสถานศึกษา ประเภทของสถานศึกษา สังกัดของสถานศึกษา ตำแหน่งที่ตั้งของสถานศึกษา ระดับการขาดแคลนครู ร้อยละงบประมาณของรัฐที่จัดสรรให้สถานศึกษา ทรัพยากรการเรียน สัดส่วนนักเรียนต่อครู และหลักสูตรการวัดและประเมินผล

โมเดลเชิงเส้นตรงแบบลดหลั่น (hierarchical linear model: HLM) หมายถึง รูปแบบการวิเคราะห์ข้อมูลที่มีตัวแปรอิสระหรือตัวแปรทำนายมากกว่าหนึ่งตัว ที่จัดลักษณะของตัวแปรได้เป็นระดับชั้นอย่างน้อย 2 ระดับ และมีตัวแปรตามหรือตัวแปรเกณฑ์อยู่ในระดับที่ 1 โดยผลการวิเคราะห์ในระดับต่ำกว่าสามารถนำมาใช้ในการวิเคราะห์ในระดับที่สูงขึ้น ทำให้สามารถอธิบายอิทธิพลที่ส่งผลของตัวแปรที่อยู่ในระดับต่างกันได้

คะแนนมูลค่าเพิ่มของสถานศึกษา หมายถึง ผลต่างระหว่างคะแนนสเกลของการประเมินผลการรู้เรื่องวิทยาศาสตร์โดยเฉลี่ยของแต่ละสถานศึกษาที่ทำนายได้จากการควบคุมอิทธิพลของตัวแปรระดับนักเรียนและระดับสถานศึกษากับคะแนนสเกลของการประเมินผลการรู้เรื่องวิทยาศาสตร์โดยเฉลี่ยของสถานศึกษาที่สังเกตได้จริง ซึ่งจากผลการวิเคราะห์มูลค่าเพิ่มจะเป็นค่าส่วนที่เหลือ (residual) ของจุดตัดแกน (intercept) จากการประมาณค่าแบบเบส์ (Baye's Estimation) ด้วยหลักการวิเคราะห์ถดถอยพหุระดับจากโปรแกรมการวิเคราะห์เชิงเส้นระดับลดหลั่น

ประสิทธิภาพมูลค่าเพิ่มการจัดการศึกษาวิทยาศาสตร์ หมายถึง ค่าสัมประสิทธิ์การทำนาย (R^2) ที่ได้มาจากโมเดลการประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ซึ่งเป็นโมเดลการวิเคราะห์พหุระดับที่มีการควบคุมตัวแปรระดับนักเรียนและระดับสถานศึกษา

ประโยชน์ที่จะได้รับ

1. ทำให้การประเมินคุณภาพการจัดการศึกษามีความถูกต้อง แม่นยำ และเป็นเกิดความยุติธรรมกับผู้ที่เกี่ยวข้องทุกฝ่าย เพราะคะแนนที่ได้จากการพิจารณาการตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันและแบบสอบที่ตรวจพบการทำหน้าที่ต่างกันจะสามารถช่วยลดคะแนนเฟ้อที่เกิดขึ้นและผลการประเมินคุณภาพการจัดการศึกษายังเป็นประโยชน์ต่อการวางแผนการจัดการเพื่อส่งเสริมคุณภาพการจัดการศึกษาทั้งระดับเขตการศึกษา โรงเรียนและนักเรียนให้มีประสิทธิภาพเพิ่มขึ้น

2. ได้โมเดลมูลค่าเพิ่มพหุระดับที่ใช้วัดประสิทธิภาพของสถานศึกษาโดยคำนึงถึงตัวแปรที่ส่งต่อความคลาดเคลื่อนทั้งในระดับนักเรียนและสถานศึกษา เนื่องจากตัวแปรที่มีการควบคุมสามารถนำมาลดความคลาดเคลื่อนทั้งในระดับข้อสอบและในโมเดลมูลค่าเพิ่มพหุระดับ

3. ได้โมเดลมูลค่าเพิ่มพหุระดับที่ได้จากการลดความลำเอียงของข้อสอบและแบบสอบซึ่งทำให้มั่นใจว่าคะแนนมูลค่าเพิ่มที่ได้เป็นคะแนนที่แท้จริงที่อาจจะส่งผลต่อการประเมินคุณภาพของสถานศึกษาที่มีผลต่อการจัดอันดับโรงเรียนหรือจัดอันดับคุณภาพโรงเรียน

4. หน่วยงานที่เกี่ยวข้องกับการจัดการศึกษาทั้งในภาครัฐและเอกชนสามารถใช้ผลการวิเคราะห์ข้อสอบที่คำนึงถึงการขจัดความลำเอียงของข้อสอบและแบบสอบไปใช้ตัดสินใจและกำหนดนโยบาย ตลอดจนพัฒนาแบบทดสอบให้มีคุณภาพมากยิ่งขึ้น ทั้งนี้ยังสามารถส่งเสริมและสนับสนุนหน่วยงานที่เกี่ยวข้องนำเทคนิควิธีการตรวจสอบความลำเอียงของข้อสอบไปในการจัดการศึกษาให้มีคุณภาพมากขึ้น

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 2
แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในส่วนนี้ผู้วิจัยนำเสนอ แบ่งออกเป็น 5 ตอน ตอนที่หนึ่ง คือ มโนทัศน์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ เพื่อเป็นการทำความเข้าใจความเป็นมาของการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ ประเภทการทำหน้าที่ต่างกันของข้อสอบ รูปแบบการทำหน้าที่ต่างกันของข้อสอบในลักษณะต่างๆ และงานวิจัยที่เกี่ยวข้อง ตอนที่สอง คือ มโนทัศน์การทำหน้าที่ต่างกันเป็นขั้น เพื่ออธิบายแนวคิดและความสำคัญของวิธีการที่ใช้ในข้อสอบที่มีการตรวจให้คะแนนพหุวิภาคและงานวิจัยที่เกี่ยวข้อง ตอนที่สาม คือ รายละเอียดโครงการประเมินผลนักเรียนนานาชาติ (PISA) และงานวิจัยที่เกี่ยวข้อง ตอนที่สี่ คือ แนวคิดเกี่ยวกับมูลค่าเพิ่มของการจัดการศึกษาและการวิเคราะห์พหุระดับและงานวิจัยที่เกี่ยวข้อง และตอนสุดท้าย คือ กรอบแนวคิดในการวิจัย โดยมีรายละเอียดดังต่อไปนี้

ตอนที่ 1 มโนทัศน์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ

ในตอนที่หนึ่งผู้วิจัยแบ่งการนำเสนอออกเป็น 4 ส่วน ได้แก่ การทำหน้าที่ต่างกันของข้อสอบ (differential item functioning: DIF) ประเภทการทำหน้าที่ต่างกันของข้อสอบการทำหน้าที่ต่างกันของแบบสอบ (differential test functioning: DTF) และงานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ โดยมีรายละเอียดดังต่อไปนี้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

1.1 การทำหน้าที่ต่างกันของข้อสอบ (differential item functioning: DIF)

การทำหน้าที่ต่างกันของข้อสอบ (differential item Ffnctioning: DIF) เป็นการศึกษาความตรงของแบบสอบและความยุติธรรมของแบบสอบ (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999; Camilli, 2006 cited in Penfield, 2010; Perrone, 2006) ซึ่งในแบบสอบที่ต้องการวัดความสามารถทางด้านใดด้านหนึ่งนั้นจำเป็นต้องมีความตรงสำหรับนักเรียนที่มีความแตกต่างกันหลากหลายกลุ่ม เช่น เชื้อชาติ เพศ และสถานะทางด้านการศึกษา (Bao & Hendrickson, 2009; Barnes & Wells, 2009; Gallagher & Kaufman, 2005; Walker, Zhang, & Surber, 2008; Xi, 2010) ทั้งนี้ในการศึกษาการทำหน้าที่ต่างกันของข้อสอบ มีการศึกษามาเป็นระยะเวลานานแล้ว โดยเริ่มต้นในปี ค.ศ.1905 โดย Binet และคณะ มีการแก้ไขและพิจารณาความไม่ยุติธรรมของข้อสอบ ต่อมาในปี ค.ศ.1912 Weintrob และ Weintrob มีการศึกษาเรื่องเชื้อชาติ

และนำตัวแปรเชื้อชาติเข้ามาศึกษาในความยุติธรรมของแบบสอบ ภายหลังได้กลายเป็นประเด็นสำคัญทางการศึกษาในประเทศสหรัฐอเมริกา (Binet & Simon, 1916 cited in McNamara & Roever, 2004; ชลี ภัทรพิชญธรรม, 2553) และในปี ค.ศ.1914 Stem เป็นคนแรกที่ศึกษาเรื่องความยุติธรรมของแบบสอบ โดยคัดเลือกข้อสอบที่ยากมากหรือง่ายมากสำหรับผู้เข้าสอบที่มาจากครอบครัวที่สถานภาพทางเศรษฐกิจสูงและผู้สอบที่มีฐานะทางเศรษฐกิจต่ำ จนมาถึงช่วงต้นของคริสต์วรรษที่ 1960 เป็นการศึกษาความยุติธรรมของข้อสอบที่มีต่อผู้สอบกลุ่มต่างๆ เป็นผลมาจากความรู้สึกไม่เท่าเทียมกันในการทดสอบความสามารถทางสมองของนักเรียนในการสอบคัดเลือกต่างๆ อันเป็นผลมาจากความแตกต่างทางด้านวัฒนธรรม และด้านเชื้อชาติ แต่การศึกษาคุณภาพความยุติธรรมของข้อสอบหรือแบบสอบเริ่มมีการศึกษาอย่างจริงจังในช่วงปลายทศวรรษของปี ค.ศ. 1960 (ศิริชัย กาญจนวาสี, 2550ก) โดยมีการเสนอวิธีการต่างๆ เพื่อตรวจสอบความลำเอียงของข้อสอบ (Item bias) ความลำเอียงของแบบสอบ (test bias) และความลำเอียงในการคัดเลือก (selection bias) โดยมีการให้นิยามความลำเอียงว่าเป็น ความคลาดเคลื่อนอย่างเป็นระบบ (systematic error) ที่เกิดขึ้นจาก การวัดความพยายามของการตรวจสอบความลำเอียงดังกล่าวดำเนินไปเพื่อจำแนกข้อสอบที่ทำหน้าที่ไม่เหมาะสมหรือไม่ยุติธรรมสำหรับปรับปรุง หรือตัดข้อสอบข้อนั้นออกจากแบบสอบ เป็นการขจัดข้อสอบที่ทำให้เกิดปัญหาความยุติธรรมระหว่างกลุ่มข้อสอบกลุ่มต่างๆ ที่มีลักษณะบางอย่างแตกต่างกัน จนมาถึงปัจจุบัน คือต้นคริสต์ศตวรรษที่ 21 ซึ่งมีการศึกษาเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบ โดยนำตัวแปรอื่นๆ เข้ามาศึกษาเพิ่มเติมและนำเสนอวิทยาการที่ทันสมัยในการตรวจสอบข้อสอบที่มีการการตรวจให้คะแนนทวิวิภาคและข้อสอบที่มีการตรวจให้คะแนนพหุวิภาค (Kamata, 2009 อ้างถึงใน ชลี ภัทรพิชญธรรม, 2553; ศิริชัย กาญจนวาสี, 2550ก; Penfield, 2010)

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเป็นการเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มผู้สอบกลุ่มย่อยสองกลุ่มที่มีความสามารถระดับเดียวกัน โดยที่ผู้สอบกลุ่มหนึ่งเป็นตัวแทนกลุ่มหลักในประชากรเรียกว่า กลุ่มอ้างอิง (Reference group) ส่วนอีกกลุ่มเป็นตัวแทนกลุ่มรองในประชากรเรียกว่า กลุ่มเปรียบเทียบ (Focal group) ซึ่งเป็นกลุ่มที่สนใจที่จะศึกษาการทำหน้าที่ต่างกันของข้อสอบ ซึ่งในการทดสอบแต่ละครั้ง ผู้สอบระหว่างกลุ่มย่อยอาจมีลักษณะแตกต่างกัน เช่น เชื้อชาติ ศาสนา วัฒนธรรม ภูมิลำเนา สังคม เพศ ภาษา อายุ ประสบการณ์ เป็นต้น ผู้สอบกลุ่มย่อยอาจไม่ได้รับความยุติธรรมในการทำข้อสอบ โดยข้อสอบบางข้ออาจมีความลำเอียงเข้าข้างผู้สอบกลุ่มย่อยบางกลุ่มของผู้เข้าสอบทั้งหมด ซึ่งทำให้เกิดการได้เปรียบเทียบหรือเสียเปรียบระหว่างผู้สอบกลุ่มย่อยด้วยกัน ดังนั้นแบบสอบหรือข้อสอบจึงขาดความตรง ซึ่งสาเหตุดังกล่าวอาจมาจากแบบสอบไม่ได้วัดความสามารถเป้าหมายที่ต้องการวัด (target ability; θ) อย่างเดียวแต่ขาดการวัดความสามารถแทรกซ้อนที่ไม่ต้องการวัด (nuisance ability ; η) (Nandakumar, 1993

อ้างถึงใน วลีมาศ แซ่อึ้ง, 2543) ผู้สอบกลุ่มย่อยกลุ่มใดมีความสามารถแทรกซ้อนสูงกว่าก็มีโอกาสในการตอบข้อสอบได้ถูกต้องมากกว่า ทั้งๆ ที่ระดับความสามารถเป้าหมายที่ต้องการวัดเท่ากัน จึงมีผลทำให้ข้อสอบทำหน้าที่ต่างกัน โดยมีเกณฑ์การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจำแนกได้เป็น 2 กลุ่ม (ศิริชัย กาญจนวาสี, 2550ก) คือ กลุ่มวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบ 2 ค่า หรือ ทวิวิภาค (dichotomous DIF methods) คือการให้คะแนนแบบ 0-1 และกลุ่มวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบหลายค่า หรือ พหุวิภาค (Polytomous DIF methods) คือมีการให้คะแนนมากกว่า 2 ค่าในข้อสอบ นอกจากนี้วิธีการตรวจสอบการทำหน้าที่ต่างกันทั้งสองแบบนี้ยังสามารถจำแนกได้ 2 มิติ ได้แก่ มิติลักษณะของตัวแปรเกณฑ์ซึ่งแบ่งเป็นกลุ่มที่ใช้คะแนนสังเกตได้ (observed score) และกลุ่มวิธีที่ใช้คะแนนที่สังเกตได้หรือคะแนนของตัวแปรแฝง (latent variable) และมิติลักษณะของสถิติวิเคราะห์ ซึ่งแบ่งเป็นกลุ่มที่ใช้พาราเมทริก (parametric approach) และกลุ่มวิธีที่ใช้สถิตินันพาราเมทริก (non-parametric approach) ซึ่งมีรายละเอียดดังต่อไปนี้

1. วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนทวิวิภาค (dichotomous DIF methods) ได้แก่ 1) กลุ่มที่ใช้คะแนนที่สังเกตได้ วิธีนี้มักวิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิม (CTT) หรือกลุ่มที่ไม่ใช้ทฤษฎีการตอบสนองข้อสอบ (non-IRT approach) โดยใช้คะแนนรวมของผู้สอบเป็นเกณฑ์การจับคู่กลุ่มผู้สอบ วิธีการตรวจสอบที่สำคัญในกลุ่มนี้ได้แก่ การวิเคราะห์ความแปรปรวน (ANOVA) การวิเคราะห์ถดถอยโลจิสติก (logistics regression) วิธีแมนเทล-แฮนส์เซล (mantel-heanszel: MH) วิธีดัชนีมาตรฐาน (standardization : STND) 2) กลุ่มวิธีที่ใช้คุณลักษณะหรือตัวแปรแฝง วิธีในกลุ่มนี้ใช้คุณลักษณะหรือตัวแปรแฝงซึ่งวิเคราะห์บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบ (IRT) สำหรับเกณฑ์การจัดคู่กลุ่มผู้สอบ วิธีการตรวจสอบที่สำคัญในกลุ่มนี้ ได้แก่ วิธีวัดพื้นที่ความแตกต่างระหว่างโค้งการตอบสนองข้อสอบ (IRT-D^2) วิธีไค-สแควร์ของลอร์ด (lord's χ^2) วิธีอัตราส่วนไลค์ลิฮูดทั่วไป (general IRT likelihood ratio) วิธีอัตราส่วนไลค์ลิฮูดลอกลิเนียร์ (loglinear IRT likelihood ratio) และวิธีซิปเทสท์ (SIBTEST)

2. วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบให้คะแนนพหุวิภาค (Polytomous DIF methods) ได้แก่ 2.1) กลุ่มวิธีที่ใช้คะแนนสังเกตได้ ประกอบไปด้วยวิธีการวิเคราะห์ความแปรปรวน (ANOVA) วิธีการวิเคราะห์การถดถอยโลจิสติกพหุวิภาค (polytomous logistic regression) วิธีดัชนีมาตรฐานพหุวิภาค (polytomous standardization) วิธีแมนเทล-แฮนส์เซลทั่วไป (general mantel haenzel: GMH) 2.2) กลุ่มวิธีที่ใช้คุณลักษณะแฝง ได้แก่ วิธีอัตราส่วนไลค์ลิฮูดในรูปทั่วไป (general IRT likelihood ratio) วิธีการให้คะแนนบางส่วน (partial credit model: PCM) วิธีซิปเทสท์พหุวิภาค (polytomous SIBTEST) และวิธีการให้คะแนน

บางส่วนทั่วไป (generalized partial credit model: GPCM) จากเกณฑ์การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบดังกล่าวสามารถสรุปได้รายละเอียดตามตารางที่ 1

ตารางที่ 1 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

ประเภทและตัวแปรเกณฑ์	รูปแบบพาราเมทริก (parametric form)	รูปแบบนันพาราเมทริก (nonparametric form)
1.DIF แบบให้คะแนนทวิวิภาค 1.1 คะแนนที่สังเกตได้ (observed score)	ANOVA (Cleary & Hilton, 1968) logistic regression (Swaninathan & Rogers,1990)	TID (Cleary & Hilton, 1968; Angoff & Ford, 1973) MH (Holland & Thayer, 1988, 1989) STND (Dorans & Kulick, 1986)
1.2 คุณลักษณะ/ตัวแปรแฝง (latent variable)	IRT-D^2 (Linn et al., 1981; shepard et al., 1984; Raju, 1990; Kim & Cohen, 1991) lord's χ^2 (Lord, 1980) general IRTLR (Thissen, Steinberg & Wainer, 1993) loglinear IRTLR (Thissen, Steinberg & Wainer, 1993)	SIBTEST (Shealy & Stout, 1993)
2. DIF แบบให้คะแนนพหุวิภาค 1.1 คะแนนที่สังเกตได้ (observed score)	ANOVA (Cleary & Hilton, 1968) polytomous logistic regression (Swaninathan & Rogers, 1990)	polytomous STND (Dorans & Kulick, 1986) GMH (Holland & Thayer, 1988; 1989)
1.2 คุณลักษณะ/ตัวแปรแฝง (latent variable)	general IRTLR (Thissen, Steinberg & Wainer, 1993) PCM (Master, 1982)	polytomous SIBTEST (Shealy & Stout, 1993) GPCM (Muraki, 1992; 1993)

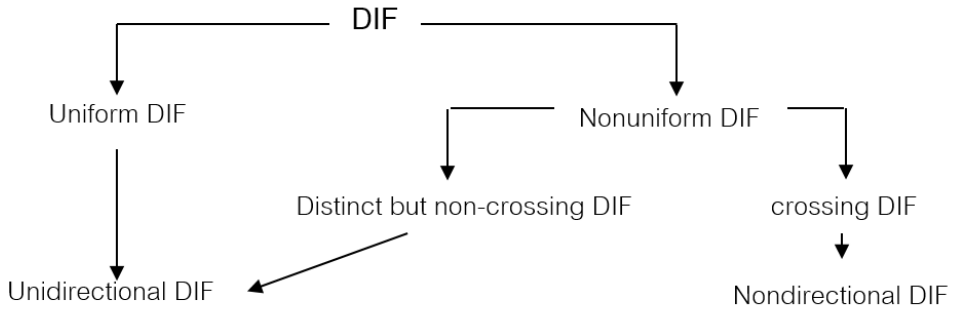
มาจาก ศิริชัย กาญจนวาสี (2550ก)

1.2 ประเภทการทำหน้าที่ต่างกันของข้อสอบ

การทำหน้าที่ต่างกันของข้อสอบเป็นการตรวจสอบการตอบสนองข้อสอบจากผู้ทดสอบ 2 กลุ่ม คือ กลุ่มอ้างอิง (Reference group: R) และกลุ่มเปรียบเทียบ (Focal group: F) โดยทั้งสองกลุ่มเป็นกลุ่มที่มีลักษณะตามประชากร เช่น เชื้อชาติ หรือ เพศ เป็นต้น Gideon (1982) ได้จำแนกลักษณะการทำหน้าที่ต่างกันของข้อสอบที่มีการตรวจให้คะแนนทวิวิภาค โดยใช้ log-linear model และ logit model ภายใต้ตารางเลขจร (contingency table) แบบสามมิติ แบ่งออกเป็น

2 ประเภท คือ ข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป (uniform DIF) หมายถึงข้อสอบที่ทำให้ผู้สอบกลุ่มหนึ่งมีโอกาสในการตอบข้อสอบถูกมากกว่าผู้สอบอีกกลุ่มหนึ่งอย่างสม่ำเสมอกันในทุกระดับความสามารถ ข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป (nonuniform DIF) หมายถึง ข้อสอบที่ทำให้ผู้สอบกลุ่มหนึ่งมีโอกาสในการตอบข้อสอบถูกมากกว่าผู้สอบอีกกลุ่มหนึ่งอย่างไม่สม่ำเสมอกัน ในทุกระดับความสามารถ และสำหรับข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปแล้วสามารถที่แยกโค้งลักษณะข้อสอบออกได้อีก 2 ประเภท คือ โค้งที่ตัดกันในข้อสอบ (crossing ICCs) และโค้งที่ต่างกันแต่ไม่ตัดกันในข้อสอบ (distinct but noncrossing DIF)

นอกจากนี้ Swaminathan และ Rogers (1990) ได้จำแนกข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป (nonuniform DIF) ได้ 2 ลักษณะ ภายใต้ทฤษฎีการตอบสนองข้อสอบ เมื่อกำหนดระดับความสามารถในช่วง -3 ถึง +3 พบว่า ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป โดยมีปฏิสัมพันธ์ไม่เป็นลำดับ (disordinal interaction) การทำหน้าที่ต่างกันสำหรับกลุ่มผู้สอบซึ่งเกิดขึ้น เมื่อโค้งลักษณะข้อสอบ (ICCs) ตัดกันหรือเรียกว่าข้อสอบทำหน้าที่ต่างกันแบบไม่มีทิศทาง (nondirectional DIF) และข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปแบบมีปฏิสัมพันธ์เป็นลำดับ (ordinal interaction) เป็นการทำหน้าที่ต่างกันสำหรับกลุ่มผู้สอบซึ่งเกิดขึ้นเมื่อโค้งลักษณะข้อสอบ (ICCs) ตัดกันไม่สม่ำเสมอ แต่ไม่ตัดกันภายขอบเขตที่ถูกสังเกตของคุณลักษณะแฝง (latent trait) อาจเรียกข้อสอบลักษณะนี้ว่า ข้อสอบทำหน้าที่ต่างกันแบบมีทิศทางเดียว (unidirectional DIF) Shealy และ Stout (1993 cited in Li & Stout, 1996) ได้กล่าวเพิ่มเติมว่า ข้อสอบทำหน้าที่ต่างกันแบบมีทิศทางเดียว (unidirectional DIF) นั้นมีลักษณะทั่วไปมากกว่าข้อสอบที่ทำหน้าที่ต่างกัน แบบเอกรูป (uniform DIF) การทำหน้าที่ตัดกันในข้อสอบเกิดจากโค้งลักษณะข้อสอบตัดกันส่วนในการทำหน้าที่ต่างกันแบบมีทิศทางเดียว (unidirectional DIF) โค้งลักษณะข้อสอบไม่ตัดกัน (non-crossing ICCs) ดังนั้นการจำแนกลักษณะการทำหน้าที่ต่างกันของข้อสอบ (DIF) ตามทฤษฎีการตอบสนองข้อสอบสำหรับข้อสอบที่มีการตรวจให้คะแนน 2 ค่า แสดงได้ตามรูปภาพที่ 1



รูปภาพที่ 1 แสดงลักษณะการทำหน้าที่ต่างกันของข้อสอบ (DIF)

1.3 การทำหน้าต่างกันของแบบสอบ (differential test functioning: DTF)

การวัดการทำหน้าที่ต่างกันของแบบสอบสามารถให้ค่าสารสนเทศที่มีค่าต่อกระบวนการตรวจสอบความตรงของแบบสอบหลายลักษณะ ประการแรก คือ การวัดการทำหน้าที่ต่างกันของแบบสอบให้ค่าสารสนเทศที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบทั้งหมดโดยการรวมทุกข้อสอบในแบบสอบ ประการที่สอง คือ เนื่องจากการประมาณค่าผลการทำหน้าที่ต่างกันของข้อสอบโดยใช้ตัวแปรที่มีการแยกเป็นชั้น ภายในแบบสอบส่งผลกระทบโดยตรงการเกิดการทำหน้าที่ต่างกันของข้อสอบภายในข้อสอบข้ออื่นๆ ในตัวแปรแยกที่มีการแยกเป็นชั้น ซึ่งการวัดการทำหน้าที่ต่างกันของแบบสอบสามารถให้ดัชนีผลกระทบที่เป็นไปได้ของการแยกเป็นชั้นที่อยู่ภายในผลการทำหน้าที่ต่างกันของข้อสอบ ทั้งนี้ระดับการทำหน้าที่ต่างของแบบสอบขนาดใหญ่มีผลกระทบที่เป็นไปได้ของการประมาณค่าผลการทำหน้าที่ต่างกันของข้อสอบมากกว่าการทำหน้าที่ต่างกันในแบบสอบขนาดเล็ก และประการสุดท้าย คือ การวัดพื้นที่ที่คิดเครื่องหมายของการทำหน้าที่ต่างกันของแบบสอบเป็นองค์ประกอบสำคัญของการวิเคราะห์การทำหน้าที่ต่างกันเป็นกลุ่มก้อน (differential bundle functioning: DBF) ในขณะที่ขนาดของการทำหน้าที่ต่างกันของแบบสอบเป็นกลุ่มก้อนสามารถใช้ระบุสาเหตุการเกิดการทำหน้าที่ต่างกันของข้อสอบได้ (Douglas, Roussos, & Stuout, 1996; Gierl, Bisanz, Bisanz, & Boughton, 2003; Gierl, Bisanz, Bisanz, Boughton, & Khaliq, 2001; Oshima, Raju, Flowers, & Slinde, 1998 cited in Penfield & Algina, 2006)

การวัดการทำหน้าที่ต่างกันของแบบสอบ จำแนกได้เป็น แบบสอบที่ข้อสอบมีการตรวจให้คะแนนทวิวิภาค แบบสอบที่ข้อสอบมีการตรวจให้คะแนนพหุวิภาคและแบบสอบที่ข้อสอบมีการตรวจให้คะแนนแบบผสมผสาน (mixed format test) โมเดลที่มีการวัดการทำหน้าที่ต่างกันของแบบสอบของข้อสอบที่มีการตรวจให้คะแนนทวิวิภาค แบ่งได้เป็น 3 แนวคิด คือ แนวคิดที่หนึ่ง Shealy และ Stout (Shealy & Stout, 1993; Raju, van der Linden, & Fleer, 1995; Dougla

et al., 1996; Gierl et al., 2001; Gierl & Khaliq, 2001; Bolt, 2000; Nandakumar, 1993 cited in Penfield & Algina, 2006) เสนอกระบวนการตรวจสอบโดยใช้ SIBTEST ซึ่งให้ผลตัวประมาณค่าของความแตกต่างระหว่างกลุ่มในเงื่อนไขค่าเฉลี่ย (the average conditional between-group difference) ในความสามารถที่ถูกคาดหวังซึ่งอยู่บนเซตย่อยของข้อสอบ (s) จากข้อสอบในแบบสอบ (n) ซึ่งกระบวนการตรวจสอบแบบ SIBTEST ได้รับความนิยมกันอย่างแพร่หลายในการทดสอบการทำหน้าที่ต่างกันของข้อสอบ การทดสอบการทำหน้าที่ต่างกันของแบบสอบ และระบุการเกิดการทำหน้าที่ต่างกันของข้อสอบที่มีประสิทธิภาพ เนื่องจากวิธี SIBTEST เป็นวิธีที่คิดพื้นที่ของการทำหน้าที่ต่างกันของข้อสอบและการทำหน้าที่ต่างกันของแบบสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบสุทธิหรือแบบแยกส่วน (net or aggregated DIF) ในทุกๆ เซตย่อยของข้อสอบ แต่วิธีดังกล่าวยังไม่เหมาะสมสำหรับการทดสอบการทำหน้าที่ต่างกันของข้อสอบแบบไม่คิดพื้นที่ระดับภาพรวม (the global level of unsigned DIF) ในทุกๆ เซตย่อยของข้อสอบ

แนวคิดที่สอง กระบวนการ DFITS ซึ่งใช้การประมาณค่าความแตกต่างกำลังสองระหว่างกลุ่มที่ถูกคาดหวัง (the expected between-group squared difference) ของคะแนนจริง เนื่องจากกระบวนการ DFITS พิจารณาความแตกต่างกำลังสองของคะแนนจริง ดังนั้นตัวประมาณค่าจึงเป็นการทำหน้าที่ต่างกันของแบบสอบแบบไม่คิดพื้นที่ (an unsigned DTF) นอกจากนี้กระบวนการดังกล่าวยังสามารถประยุกต์ใช้ได้กับแบบสอบทั้งหมด ถึงแม้ว่าอาจจะไม่ไวต่อผลการทำหน้าที่ต่างกันของข้อสอบในระดับข้อสอบซึ่งอาจจะลดประสิทธิภาพสำหรับการวัดการทำหน้าที่ต่างกันของแบบสอบในระดับแบบสอบตามมา ทั้งนี้กระบวนการ DFITS จำเป็นที่จะต้องมีการประมาณค่าพารามิเตอร์ตามทฤษฎีตอบสนองข้อสอบทั้งกลุ่มอ้างอิงและกลุ่มเปรียบเทียบและขนาดของแต่ละกลุ่มต้องมีขนาดใหญ่ด้วย

แนวคิดที่สาม โมเดลอิทธิพลเชิงสุ่ม (a random effects model) และประเมินความแปรปรวนในผลการทำหน้าที่ต่างกันของข้อสอบของทุกข้อในแบบสอบ การวัดการทำหน้าที่ต่างกันของแบบสอบอยู่บนพื้นฐานการประมาณค่าความแปรปรวนผลการทำหน้าที่ต่างกันของข้อสอบซึ่งอ้างอิงไปยังการวัดการทำหน้าที่ต่างกันของแบบสอบที่อยู่บนพื้นฐานความแปรปรวน (variance-based DTF measures) โดยที่ข้อได้เปรียบของการวัดการทำหน้าที่ต่างกันของแบบสอบบนพื้นฐานความแปรปรวนเป็นการวัดการทำหน้าที่ต่างกันของแบบสอบแบบไม่คิดพื้นที่และเหมาะอย่างยิ่งที่จะบอกจำนวนของผลการทำหน้าที่ต่างกันของข้อสอบแบบภาพรวม (the global DIF effect) ซึ่งเป็นแนวคิดที่ต่างจากวิธี SIBTEST และ DFITS โดยดัชนีค่าคาดหวังของการทำหน้าที่ต่างกันของแบบสอบบนพื้นฐานความแปรปรวน (the expected value of variance-based DTF indices) สามารถระบุในแต่ละข้อสอบในแบบสอบ สามารถเปรียบเทียบสถานการณ์ของแบบสอบที่แตกต่างกันและสามารถประเมินความสัมพันธ์ในแต่ละเกณฑ์ที่กำหนดมาได้ นอกจากนี้การวัดการทำหน้าที่

ต่างกันของแบบสอบบนพื้นฐานความแปรปรวนไม่เหมาะสมกับการทดสอบตามทฤษฎีตอบสนองข้อสอบแต่จะเหมาะสมกับการทดสอบกับกลุ่มตัวอย่างมีขนาดเล็ก

โมเดลที่มีการวัดการทำหน้าที่ต่างกันของแบบสอบของข้อสอบมีการตรวจให้คะแนนพหุวิภาคสามารถใช้กระบวนทางสถิติประเภทเดียวกับที่ใช้ตรวจสอบการทำหน้าที่ต่างกันของแบบสอบของข้อสอบที่มีการตรวจให้คะแนนทวิวิภาค โดยใช้วิธี SIBTEST และ DFITS ส่วนโมเดลการวัดการทำหน้าที่ต่างกันของแบบสอบของข้อสอบมีการตรวจให้คะแนนแบบผสมผสาน (mixed format tests) (Penfield & Algina, 2006) สามารถใช้ตัวประมาณค่าของการทำหน้าที่ต่างกันของแบบสอบที่มีพื้นฐานอยู่บนความแปรปรวนของผลการทำหน้าที่ต่างกันของข้อสอบ (the DIF effect variance) โดยหลักการดังกล่าวทำให้การพัฒนาตัวประมาณค่าการทำหน้าที่ต่างกันของแบบสอบแบบทั่วไป (the generalized DTF estimators) เพื่อตรวจสอบความยุติธรรมของแบบสอบในแบบสอบที่มีรูปแบบทั้งการตรวจให้คะแนนทวิวิภาคและพหุวิภาคในฉบับเดียวกัน โดยมีรายละเอียดดังนี้

1.3.1 ตัวประมาณค่าการทำหน้าที่ต่างกันของข้อสอบทั่วไป(the generalized DIF estimators)

Penfield และ Algina (2003) ได้เสนอตัวประมาณค่าอัตราส่วนเลขออดทั่วไปสะสม (an estimator of the cumulative common odds ratio) สำหรับตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการตรวจให้คะแนนพหุวิภาค ตัวประมาณค่าลัว-อเกรสตี (liu-agresti) ของอัตราส่วนเลขออดทั่วไปสะสมเขียนแทนด้วย $\hat{\alpha}_{LAi}$ โดยมี $\log(\hat{\alpha}_{LAi})$ เป็นตัวประมาณค่าผลการทำหน้าที่ต่างกันของข้อสอบทั่วไป (the generalized DIF effect) นอกจากนี้วิธีการตรวจสอบแบบสอบที่มีการให้คะแนนแบบผสมผสาน (mixed format tests) โดยใช้ค่าความแปรปรวนของผลการทำหน้าที่ต่างกันของข้อสอบทั่วไป (the variance of the generalized DIF effect) แทนด้วย v^2 โดยที่ตัวประมาณค่าแบ่งออกได้เป็นสองประเภท คือแบบไม่ถ่วงน้ำหนัก (unweighted estimators) อธิบายได้ตามสมการ 2.1 และแบบถ่วงน้ำหนัก (weighted estimators) อธิบายได้ตามสมการ 2.2 ดังนี้

ประเภทไม่ถ่วงน้ำหนัก (unweighted estimators)

$$\hat{\upsilon}^2 = \frac{\sum_{i=1}^{n}[\log(\hat{\alpha}_{LAi}) - \hat{\mu}]^2 - \sum_{i=1}^{n} s_i^2}{n} \quad \text{.....(2.1)}$$

ประเภทถ่วงน้ำหนัก (weighted estimators)

$$\hat{\upsilon}_w^2 = \frac{\sum_{i=1}^{n} [\log(\hat{\alpha}_{LAi}) - \hat{\mu}]^2 - \sum_{i=1}^{n} w_i}{\sum_{i=1}^{n} w_i^2} \quad (2.2)$$

โดยที่

s_i^2 หมายถึง ความแปรปรวนที่ถูกประมาณค่าของ $\log(\hat{\alpha}_{LAi})$ สำหรับข้อสอบข้อที่ i และ $w_i = s_i^{-2}$

$\hat{\mu}$ หมายถึง ค่าเฉลี่ยของ $\log(\hat{\alpha}_{LAi})$ ทุกข้อสอบในแบบสอบ

1.4 งานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ

จากการศึกษางานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ พบว่า การศึกษาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบตั้งแต่อดีตจนถึงปัจจุบันมีนักวิชาการได้ศึกษาไว้หลากหลายด้าน ทั้งนี้ผู้วิจัยได้จำแนกงานวิจัยที่สำคัญดังกล่าว ออกเป็น 5 กลุ่ม ได้แก่ กลุ่มที่ 1 ศึกษาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ (รัชนก ยี่สุ่นศรี, 2544) กลุ่มที่ 2 ศึกษาประสิทธิภาพการทำหน้าที่ต่างกันของข้อสอบตามขนาดอิทธิพล (ธเกียรติกมล ทองงอก, โชติกา ภาษีผล, และ ศิริชัย กาญจวาสี, 2556; Kim et al., 2007) กลุ่มที่ 3 ศึกษาลักษณะการทำหน้าที่ต่างกันของข้อสอบ (Finch & French, 2007; Penfield & GÜler, 2009; Taylor & Lee, 2012) กลุ่มที่ 4 ศึกษาการทำหน้าที่ต่างกันของตัวลวง (Penfield, 2010c; Penfield, 2010d) และกลุ่มที่ 5 ศึกษาการทำหน้าที่ต่างกันของข้อสอบในข้อสอบที่มีการตรวจคะแนนพหุวิภาค (Penfield, 2007a; Carvajal & Skorupski, 2010)

กุล่มที่ศึกษาเกี่ยวกับการการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ ได้แก่ รัชนก ยี่สุ่นศรี (2544) ซึ่งวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบด้วยกระบวนการ ดี เอฟ ไอ ที สำหรับแบบสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษา วิชาภาษาอังกฤษและวิชาคณิตศาสตร์ พบว่า แบบสอบวิชาภาษาอังกฤษทำหน้าที่ต่างกันตามเพศและสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียนที่จบการศึกษาของผู้สอบ ส่วนแบบสอบวิชาคณิตศาสตร์ทำหน้าต่างกันตามเพศของผู้สอบและผลการวิเคราะห์การทำหน้าที่ต่างกันที่ระดับข้อสอบ พบว่า ข้อสอบทำหน้าที่ต่างกันตามเพศของผู้สอบมากที่สุดสอดคล้องกันทั้งสองวิชา นอกจากนี้ผลการเปรียบเทียบคุณภาพของแบบสอบฉบับก่อนและหลังตัดข้อสอบที่พบ DIF ทั้งในกรณีตัดทุกข้อและบางข้อทั้ง 2 วิชาได้ผลสอดคล้องกัน โดยที่แบบสอบฉบับก่อนและหลังตัดข้อสอบมีค่าความตรงเชิงโครงสร้างของแบบสอบไม่แตกต่างกัน แบบสอบฉบับหลังตัดข้อสอบที่พบ DIF ส่วนใหญ่มีค่าความเที่ยงลดลง และ

แบบสอบฉบับหลังตัดข้อสอบที่ DIF ส่วนใหญ่มีค่าฟังก์ชันสารสนเทศของแบบสอบมีมากขึ้น ส่วนการวิเคราะห์ความสัมพันธ์ระหว่างตำแหน่งของคะแนนรวมของผู้สอบก่อนและหลังตัดข้อสอบที่พบ DIF ทั้งในกรณีตัดทุกข้อและบางข้อ พบว่า ในทุกกรณีมีความสัมพันธ์ในทางบวกซึ่งกันและกันอย่างมีนัยสำคัญ แบบสอบวิชาภาษาอังกฤษมีค่าสัมประสิทธิ์สหสัมพันธ์สเปียร์แมนแรงค์อยู่ระหว่าง 0.83-0.99 ส่วนในแบบสอบวิชาคณิตศาสตร์มีค่าสัมประสิทธิ์สหสัมพันธ์สเปียร์แมนแรงค์อยู่ระหว่าง 0.87-0.98

กลุ่มที่ศึกษาประสิทธิภาพการทำหน้าที่ต่างกันของข้อสอบตามขนาดอิทธิพล ได้แก่ ธเกียรติ-กมล ทองงอก, โชติกา ภาษีผล, และ ศิริชัย กาญจวาสี (2556) ซึ่งศึกษาประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในวิธีถดถอยโลจิสติกโดยใช้เกณฑ์ขนาดอิทธิพล 2 วิธี สำหรับข้อสอบที่มีรูปแบบการตรวจให้คะแนนแบบทวิวิภาค โดยใช้ข้อมูลจากการจำลองข้อมูลและข้อมูลเชิงประจักษ์ พบว่า วิธีถดถอยโลจิสติกโดยการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin และ Gierl มีอัตราความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูงกว่าเกณฑ์ Zumbo และ Thomas นอกจากนี้ข้อสอบที่ทำหน้าที่ต่างกันแบอเนกรูป แบบสอบที่มีจำนวนข้อสอบทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 20 มีอัตราความถูกต้องจากการวัดขนาดอิทธิพลทั้ง 2 เกณฑ์สูงกว่าในแบบสอบที่มีจำนวนข้อสอบที่ทำหน้าที่ต่างกัน ทั้งฉบับคิดเป็นร้อยละ 10 และเมื่อขนาดอิทธิพลของข้อสอบที่การทำหน้าที่ต่างกันเพิ่มขึ้นมีผลทำให้อัตราความถูกต้องจากการวัดขนาดอิทธิพลทั้ง 2 เกณฑ์เพิ่มขึ้นภายใต้เกือบทุกเงื่อนไข และผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในข้อมูลเชิงประจักษ์พบว่า ขนาดอิทธิพลตามเกณฑ์ Jodoin และ Gierl ให้อัตราความถูกต้องสูงกว่าและมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าเกณฑ์ของ Zumbo และ Thomas เมื่อข้อมูลเชิงประจักษ์มีประชากรขนาดใหญ่สามารถตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน ด้วยการทดสอบระดับนัยสำคัญอย่างมีนัยสำคัญส่งผลให้ความคลาดเคลื่อนประเภทที่ 1 มีแนวโน้มสูงขึ้น

นอกจาก ธเกียรติกมล ทองงอก, โชติกา ภาษีผล, และ ศิริชัย กาญจวาสี (2556) แล้ว Kim และคณะ (2007) ได้ศึกษาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและขนาดอิทธิพลในข้อสอบที่มีการตรวจให้คะแนนพหุวิภาคโดยใช้ข้อมูลจากการประเมินความสามารถระดับชาติ โดยใช้วิธีการตรวจจับการทำหน้าที่ต่างกันของข้อสอบ 4 วิธี โดยที่ 2 วิธี เป็นวิธีพื้นฐานตามโมเดลทฤษฎีการตอบสนองข้อสอบ (IRT model-based method) คือวิธี likelihood ratio test for IRT และวิธี logistic regression likelihood ratio test และอีก 2 เป็นวิธีตามแนวคิดของแมนเทล-แฮนส์เซล คือ วิธี mantel test และ วิธี generalized mantel-haenszel test จากผลการวิจัยชี้ให้เห็นว่า เนื่องจากอำนาจการทดสอบทางสถิติเป็นการหน้าที่ของกลุ่มตัวอย่าง การตรวจจับการทำหน้าที่ต่างกันของข้อสอบมาจากกลุ่มข้อมูลที่มีขนาดใหญ่ไม่ได้ใช้ประโยชน์ในเชิงปฏิบัติ และ indices of

standardized และ observed-score indices of standardized มีผลต่อข้อสอบที่มีการตรวจให้คะแนนพหุวิภาคซึ่งได้มาจากเปรียบเทียบ ขนาดอิทธิพลจากค่า R^2 ตาม logistic regression

นอกจากนี้ กลุ่มที่ศึกษาลักษณะการทำหน้าที่ต่างกันของข้อสอบ ได้แก่ Holmes, Finch, และ French (2007) ได้ตรวจสอบวิธีการทำหน้าที่ต่างกันแบบตัดกันของข้อสอบ โดยเปรียบเทียบจากวิธี 4 วิธี ได้แก่ วิธีซิบเทสท์ (SIBTEST) วิธีการถดถอยโลจิสติก (LR) วิธีการทดสอบอัตราส่วนไลค์ลิฮู้ดตามทฤษฎีการตอบสนองข้อสอบ (IRTLR) และการวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA) โดยองค์ประกอบที่ศึกษาประกอบไปด้วย ขนาดกลุ่มตัวอย่าง (sample size) ความสามารถที่แตกต่างระหว่างกลุ่ม (ability differences between groups) ร้อยละของการเกิดการทำหน้าที่ต่างกันของข้อสอบ (percentage of DIF) โดยศึกษาการจำลองข้อมูล ซึ่งผลการวิจัยพบว่า ทุกวิธีสามารถควบคุมความคลาดเคลื่อนประเภทที่หนึ่ง (Type I error) แต่วิธีซิบเทสท์ (SIBTEST) มีอำนาจในการตรวจสอบสูงสุด

นอกจาก Holmes, Finch, และ French (2007) แล้ว Penfield และ GÜler (2009) ได้เปรียบเทียบวิธีการถดถอยโลจิสติก (LR) และวิธีแบบตารางเลขจรณ์ คือ วิธีแมนเทล- แฮนส์เซล (MH) และวิธีเบรสโรล-เดย์ (BD) สำหรับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีลักษณะแบบเอกรูป (Uniform DIF) และอเนกรูป (Nonuniform DIF) โดยใช้การศึกษาจากการจำลองข้อมูล และเปรียบเทียบความคลาดเคลื่อนประเภทที่หนึ่ง (Type I error) และอำนาจของกฎผสานการตัดสินใจ (power of a combined decision rule) ซึ่งเป็นกฎที่เกิดจากการรวมระหว่างวิธีเบรสโรล-เดย์ (BD) และวิธีแมนเทล- แฮนส์เซล (MH) และเปรียบเทียบกับวิธีการถดถอยโลจิสติก (LR) ซึ่งผลการวิจัย พบว่า อัตราความคลาดเคลื่อนประเภทที่หนึ่ง (Type I error rate) ตามกฎผสมการตัดสินใจ (a combined decision rule) อยู่ในระดับความคงที่ที่ต่ำในระดับแอลฟาเชิงกลุ่ม (the nominal alpha level) ส่วนในวิธีการถดถอยโลจิสติก (LR) อยู่ในระดับที่สูงสำหรับเงื่อนไขที่มีการแจกแจงความสามารถไม่เท่ากัน (unequal ability distribution) นอกจากนี้อำนาจในการทดสอบ ตามกฎผสานการตัดสินใจ (power of a combined decision rule) อยู่ในระดับความคงที่ที่สูงมากกว่าวิธีการถดถอยโลจิสติก (LR) ในทุกๆ รูปแบบที่มีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

ส่วน Taylor และ Lee (2012) ได้ศึกษาการทำหน้าที่ต่างกันของข้อสอบระหว่างเพศของนักเรียนเกรด 4, 7, และ 10 ในข้อสอบวิชาการอ่านและคณิตศาสตร์ ซึ่งการทำหน้าที่ต่างกันของข้อสอบระหว่างเพศ ใช้การทดสอบผ่าน POLYSIBTEST และ Rasch พบว่า ในข้อสอบการอ่านและคณิตศาสตร์ โดยทั่วไปในข้อสอบที่มีตัวเลือกหลายตัวเลือกเข้าข้างนักเรียนชาย ในขณะที่ข้อสอบอัตนัยที่ต้องตอบแบบมีโครงสร้างเข้าข้างนักเรียนหญิง การวิเคราะห์เนื้อหาแสดงให้เห็นว่าข้อสอบประเภทการอ่านวัดความเข้าใจในเนื้อหาหรือความหมายที่มีนัยยะ นักเรียนชายมักจะทำ

ข้อสอบได้จากข้อสอบระบุการตีความหมายอย่างมีเหตุผลและการวิเคราะห์เนื้อหาของสิ่งที่เรียน ส่วนนักเรียนหญิงมักจะทำข้อสอบได้จากข้อสอบที่ต้องใช้การตีความเฉพาะบุคคล นอกจากนี้การวิเคราะห์เนื้อหาของข้อสอบวิชาคณิตศาสตร์แสดงให้เห็นว่าข้อสอบเข้าข้างนักเรียนชายที่มีการทดสอบในเนื้อหาวิชาเรขาคณิต ความน่าจะเป็น และพีชคณิต ส่วนข้อสอบวิชาคณิตศาสตร์ที่เข้าข้างนักเรียนหญิงเป็นการตีความเชิงสถิติ การแก้ปัญหาหลากหลายขั้นตอน และการให้เหตุผลเชิงคณิตศาสตร์

กลุ่มที่ศึกษาการทำหน้าที่ต่างกันของตัวลวง ได้แก่ Penfield (2010c) ซึ่งศึกษาโมเดลผลการทำหน้าที่ต่างกันของข้อสอบโดยใช้ผลความไม่แปรระดับของตัวลวง (Distractor-Level Invariance) ซึ่งเป็นนัยยะสำคัญสำหรับการทำความเข้าใจของสาเหตุการทำหน้าที่ต่างกันของข้อสอบ (DIF) ซึ่งแสดงว่าความไม่แปรเปลี่ยนการวัดข้ามทุกๆ ตัวเลือกในข้อสอบประเภทหลายตัวเลือก (multiple-choice item) สามารถใช้ norminal response model (NRM) ที่ประกอบด้วยผลการทำหน้าที่ต่างกันของตัวลวง สำหรับตัวลวงแต่ละตัวทำให้เข้าใจแนวคิดเพื่อที่จะพิจารณาผลการทำหน้าที่ต่างกันของตัวลวง (DDF) ที่ถูกกำหนดโดยผลของผลการทำหน้าที่ต่างกันของตัวลวง (DDF) ลำดับที่ j โดยมีเงื่อนไขการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีลักษณะแบบเอกรูป (uniform DIF) อเนกรูป (nonuniform DIF) และการตรวจสอบการทำหน้าที่ตัดกันในข้อสอบ ผลการวิจัยชี้ให้เห็นศักยภาพของลักษณะระดับข้อสอบ (item-level properties) ที่นำไปสู่ทั้ง 3 ลักษณะ อีกทั้งข้อค้นพบอาจจะให้มูลเหตุของรูปแบบการทำหน้าที่ต่างกันของข้อสอบที่จะช่วยวิเคราะห์สาเหตุการทำหน้าที่ต่างกันของข้อสอบในลักษณะเฉพาะได้

และในปีต่อมา Penfield (2010d) ยังได้ศึกษารูปแบบและขนาดของผลการทำหน้าที่ต่างกันของข้อสอบ ในข้อสอบประเภทหลายตัวเลือก ที่ถูกกำหนดโดยผลของความไม่แปรเปลี่ยนระดับตัวลวง ในโมเดล NRM จากผลการวิจัยชี้ให้เห็นว่า 1) การมีผลการทำหน้าที่ต่างกันของตัวลวง (a differential distractor functioning : DDF) หนึ่งตัว หรือมากกว่าศูนย์ (nonzero) ทำให้ทราบผลการทำหน้าที่ต่างกันของตัวลวงประเภทไม่เป็นศูนย์ (nonzero DIF effect) 2) ขนาดของผลการทำหน้าที่ต่างกันของตัวลวง (DDF) สร้างจากขอบเขตบนไปจนถึงขนาดของผลจากการทำหน้าที่ต่างกันของข้อสอบ 3) ผลการทำหน้าที่ต่างกันของตัวลวงขนาดใหญ่ (a large DDF effect) ไม่ได้เป็นตัวที่บอกถึงผลการทำหน้าที่ต่างกันของข้อสอบขนาดใหญ่ (a large DIF effect) 4) ภายใต้เงื่อนไขของผลการทำหน้าที่ต่างกันของตัวลวงที่คงที่เป็นผลจากการทำหน้าที่ต่างกันของข้อสอบที่เป็นอิสระกับคุณลักษณะข้อสอบ แต่ภายใต้เงื่อนไขของผลการทำหน้าที่ต่างกันของตัวลวงที่มีความแปรเปลี่ยนเป็นผลจากขนาดผลการทำหน้าที่ต่างกันของข้อสอบที่ขึ้นอยู่กับคุณลักษณะข้อสอบ และ 5) ถึงแม้ว่าการตรวจสอบการทำหน้าที่ต่างกันแบบตัดกันในข้อสอบสามารถเกิดได้เพียงกรณีเดียวคือผลการทำหน้าที่ต่างกันของตัวลวงที่มีลักษณะลู่ออก (divergent DDF effects) (เครื่องหมายต่างกัน)

แต่ว่าผลการทำหน้าที่ต่างกันของตัวลวงที่มีลักษณะลู่ออกก็ไม่ได้เกิดการตรวจสอบการทำหน้าที่ต่างกันแบบตัดกันในข้อสอบ (CDIF) เสมอไป

กลุ่มที่ศึกษาการทำหน้าที่ต่างกันของข้อสอบในข้อสอบที่มีการตรวจคะแนนพหุวิภาค ได้แก่ Penfield (2007a) ได้ศึกษาวิธีการสำหรับการจัดกลุ่มการทำหน้าที่ต่างกันของข้อสอบในข้อสอบที่มีการตรวจให้คะแนนพหุวิภาคของหน่วยงานให้บริการการทดสอบทางการศึกษา (ETS) มีการจัดระดับการทำหน้าที่ต่างกันในข้อสอบซึ่งมีพื้นฐานมาจากการแปลงค่าจาก อัตราส่วนเลขออดตามวิธีแมนเทล- แฮนส์เซล (mantel-haenzel common odds ratio: $\hat{\lambda}_{MH}$) โดยที่การวิจัยครั้งนี้มีจัดกลุ่มแบบแผนการทำหน้าที่ต่างของข้อสอบสำหรับข้อสอบที่มีการตรวจให้คะแนนพหุวิภาคแบ่งเป็น 2 แบบคือ แบบแผน P1 (ข้อสอบที่มีการเดา) และ แบบแผน P2 (ข้อสอบที่ไม่มีการเดา) ผลจากการศึกษาทั้ง 2 แบบสอดคล้องกับเกณฑ์จากหน่วยงานให้บริการการทดสอบทางการศึกษา ซึ่งใช้ $\hat{\lambda}_{MH}$ เป็นตัวประมาณค่าผลการทำหน้าที่ต่างกันของข้อสอบ (the DIF effect estimator) และทั้งนี้ทั้ง 2 รูปแบบที่ศึกษาสามารถนำไปประยุกต์ใช้กับข้อสอบที่มีการตรวจให้คะแนนพหุวิภาค

นอกจากนี้ นักวิชาการที่ศึกษาการทำหน้าที่ต่างกันของข้อสอบในข้อสอบที่มีการตรวจคะแนนพหุวิภาค นอกจาก Penfield (2007a) แล้ว ในปีต่อมา Carvajal และ Skorupski (2010) ยังได้ศึกษาผลของขนาดกลุ่มตัวอย่างขนาดเล็กในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในข้อสอบที่มีการตรวจให้คะแนนพหุวิภาคโดยใช้ตัวประมาณค่า liu-agresti ของอัตราส่วนเลขออดสะสม (cumulative common odds ratio) ทั้งนี้ตัวประมาณค่า liu-agresti ถูกนำเสนอโดย Penfield และ Algina ซึ่งเป็นวิธีที่ใช้ศึกษาการทำหน้าที่ต่างกันของข้อสอบของข้อสอบที่มีการตรวจให้คะแนนพหุวิภาคโดยการวิจัยครั้งนี้ศึกษากับกลุ่มตัวอย่างขนาดเล็กเพื่อวิเคราะห์ตัวประมาณค่าดังกล่าว พิจารณาจาก ขนาดอิทธิพล (effect size) ความคลาดเคลื่อนประเภทที่หนึ่ง (Type I error) และอัตราอำนาจทดสอบ (power rates) ผลการวิจัยชี้ให้เห็นว่า กลุ่มตัวอย่างที่มีขนาดเล็กกว่า 200 มีอำนาจการทดสอบน้อยมาก ซึ่งสังเกตได้จากการทดสอบนัยสำคัญทางสถิติ อย่างไรก็ตาม อัตราความคลาดเคลื่อนประเภทที่หนึ่งเข้าใกล้ในระดับเชิงกลุ่ม (nominal level) และพบว่า log odd ratio เป็นขนาดอิทธิพลที่ไม่ผลในเชิงเปรียบเทียบสำหรับข้อสอบที่มีการจำแนกสูง

จากการศึกษางานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ สามารถสรุปผลการวิเคราะห์องค์ความรู้ที่ได้จากงานวิจัยที่ศึกษารายละเอียดตามตารางที่ 2 ดังนี้

ตารางที่ 2 สรุปผลองค์ความรู้จากงานวิจัยการทำหน้าที่ต่างกันของข้อสอบ

ผู้วิจัย	วิธีการดำเนินงาน	องค์ความรู้ในการศึกษา
รัชนก ยี่สุ่นศรี (2544)	วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบด้วยกระบวนการ ดี เอฟ ไอ ที	ใช้กระบวนการ DFIT ในการวิเคราะห์ข้อสอบและแบบสอบ
ธเกียรติกมล ทองงอก, โชติกา ภาษีผล, และศิริชัย กาญจนวาสี (2556)	ศึกษาประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในวิธีถดถอยโลจิสติกโดยใช้เกณฑ์ขนาดอิทธิพล 2 วิธี สำหรับข้อสอบที่มีรูปแบบการตรวจให้คะแนนแบบทวิวิภาค	ขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl มีอัตราความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูงกว่าเกณฑ์ Zumbo and Thomas
Holmes, Finch, และ French (2007)	ศึกษาการทำหน้าที่ต่างกันแบบตัดกันของข้อสอบ	เปรียบเทียบการทำหน้าที่ต่างกันแบบตัดกันของข้อสอบจากวิธี 4 วิธี ได้แก่ SIBTEST, LR, IRTLR, และ CFA
Penfield (2007a)	จัดกลุ่มการทำหน้าที่ต่างกันในข้อสอบในข้อสอบที่มีการตรวจให้คะแนนพหุวิภาค	จัดกลุ่มโดยใช้ $\hat{\alpha}_{\text{LA}}$ เป็นตัวประมาณค่าผลการทำหน้าที่ต่างกันของข้อสอบ (the DIF effect estimator) ในข้อสอบที่มีการตรวจให้คะแนนพหุวิภาค
Kim และคณะ (2007)	ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและขนาดอิทธิพลในข้อสอบที่มีการตรวจให้คะแนนหลายค่า	ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 4 วิธี ได้แก่ likelihood ratio test for IRT, logistic regression likelihood ratio test, mantel test และ generalized mantel-haenzel test
Penfield และ GÜler (2009)	เปรียบเทียบวิธีการหน้าที่ต่างกันของข้อสอบที่มีลักษณะแบบเอกรูป (uniform DIF) และอเนกรูป (nonuniform DIF)	ใช้ 4 วิธี ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีลักษณะแบบเอกรูป (uniform DIF) และอเนกรูป (nonuniform DIF) ได้แก่ถดถอยโลจิสติก (LR), แมนเทล- แฮนส์เซล (MH) และวิธีเบรสโลล-เดย์ (BD)

ตารางที่ 2 สรุปผลองค์ความรู้จากงานวิจัยการทำหน้าที่ต่างกันของข้อสอบ (ต่อ)

ผู้วิจัย	วิธีการดำเนินงาน	องค์ความรู้ในการศึกษา
Penfield (2010c)	ศึกษาโมเดลผลการทำหน้าที่ต่างกันของข้อสอบโดยใช้ผลความไม่แปรระดับของตัวลวง (distractor-level invariance)	ผลของผลการทำหน้าที่ต่างกันของตัวลวง (DDF) เป็นพื้นฐานในการตรวจสอบการทำหน้าที่ตัดกันในข้อสอบ (CDIF) สำหรับข้อสอบที่มีการตรวจให้คะแนนพหุวิภาค
Penfield (2010d)	ศึกษารูปแบบและขนาดของผลการทำหน้าที่ต่างกันของข้อสอบ ในข้อสอบประเภทหลายตัวเลือก ที่ถูกกำหนดโดยผลของความไม่แปรเปลี่ยนระดับตัวลวง	ทราบรูปแบบและขนาดการทำหน้าที่ต่างกันของตัวลวง
Carvajal และ Skorupski (2010)	ศึกษาผลของขนาดกลุ่มตัวอย่างขนาดเล็กในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในข้อสอบที่มีการตรวจให้คะแนนพหุวิภาค	ใช้ตัวประมาณค่า liu - agresti ของอัตราส่วนเลขออดสะสม (cumulative common odds ratio) ศึกษาการทำหน้าที่ต่างกันของข้อสอบในข้อสอบมีการตรวจให้คะแนนพหุวิภาค
Taylor และ Lee (2012)	ศึกษาการทำหน้าที่ต่างกันของข้อสอบระหว่างเพศในข้อสอบวิชาการอ่านและคณิตศาสตร์	ใช้ POLYSIBTEST และ Rasch ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีทั้งข้อสอบประเภทหลายตัวเลือกและการตอบแบบมีโครงสร้าง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

การศึกษาการทำหน้าที่ต่างกันของข้อสอบ (DIF) เป็นประเด็นทางการศึกษาที่นักวัดผลทางการศึกษาควรให้ความสำคัญ ประวัติความเป็นมาในการศึกษาการทำหน้าที่ต่างกันของข้อสอบมีการศึกษามาเป็นระยะเวลานาน ซึ่งมีทั้งการศึกษาในเชิงทฤษฎีโดยอาศัยการจำลองข้อมูลและการศึกษาโดยอาศัยข้อมูลที่มีอยู่ในสภาพจริงตามบริบทต่างๆ ทั้งนี้เมื่อพิจารณาการทำหน้าที่ต่างกันทั้งในข้อสอบที่มีการตรวจให้คะแนน 2 ค่าหรือทวิวิภาค และข้อสอบที่มีการตรวจให้คะแนนหลายค่าหรือพหุวิภาค โดยมีการศึกษาในหลายตัวแปร เช่น เชื้อชาติ ศาสนา วัฒนธรรม ภูมิลำเนา สังคม เพศ ภาษา อายุ ประสบการณ์ เป็นต้น สามารถแยกวิธีที่มีการศึกษาการทำหน้าที่ต่างกันของข้อสอบ เป็นประเภท 2 ประเภทหลักๆ คือ ตามทฤษฎีการทดสอบแบบดั้งเดิม (CTT) และตามทฤษฎีการทดสอบแนวใหม่ (IRT) ทั้งที่อยู่ในรูปแบบพาราเมทริก และรูปแบบนันพาราเมทริก

นอกจากนี้ประเภทการทำหน้าที่ต่างกันของข้อสอบ ยังสามารถแบ่งออกได้เป็นแบบเอกรูป (uniform DIF) และแบบอเนกรูป (nonuniform DIF) ซึ่งทั้งสองลักษณะของข้อสอบยังสามารถแยกประเภทการทำหน้าที่ต่างกันออกเป็น 2 ทิศทางคือ แบบมีทิศทางเดียว (unidirectional DIF) เกิดขึ้นได้กับข้อสอบที่โค้งที่ต่างกันแต่ไม่ตัดกันในข้อสอบ (distinct but noncrossing DIF) และแบบไม่มีทิศทาง (nondirectional DIF) เกิดขึ้นได้กับคือ โค้งที่ตัดกันในข้อสอบ (crossing ICCs)

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง แม้ว่าจะมีการศึกษากับการทำหน้าที่ต่างกันของข้อสอบและแบบสอบมาเป็นระยะเวลานาน ส่วนใหญ่เป็นการศึกษาการทำหน้าที่ต่างกันของข้อสอบตามลักษณะการเกิดการทำหน้าที่ต่างกันของข้อสอบหรือการทำหน้าที่ต่างกันของแบบสอบโดยทั่วไป ภายหลังนี้ยังมีนักวิชาการบางท่านที่พัฒนาองค์ความรู้ทางด้านการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการศึกษาการทำหน้าที่ต่างกันของตัวลวงในข้อสอบเพื่อนำเสนอสารสนเทศที่เป็นประโยชน์ต่อวงการวิชาการทางการวัดผล ซึ่งนับว่าเป็นความพยายามที่จะพัฒนาองค์ความรู้ทางด้านการศึกษาการทำหน้าที่ต่างกันของข้อสอบให้มีความหลากหลายมากยิ่งขึ้น

อย่างไรก็ตาม จากผลการสังเคราะห์งานวิจัยยังพบอีกว่า งานวิจัยในกลุ่มที่ศึกษาการทำหน้าที่ต่างกันของแบบสอบในข้อสอบที่มีทั้งการตรวจให้คะแนนทวิวิภาคและพหุวิภาคหรือแบบผสมผสาน (mixed format test) ในปัจจุบันมีจำนวนค่อนข้างน้อย ซึ่งมีเพียงการศึกษาของ Penfield และ Algina (2006) ซึ่งศึกษาตัวประมาณค่าการทำหน้าที่ต่างกันของข้อสอบทั่วไป (the generalized DIF estimators) ดังนั้น จากเหตุผลและความสำคัญดังกล่าวทำให้ผู้วิจัยมีแนวคิดที่จะพัฒนาองค์ความรู้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและแบบสอบให้มีความกว้างขวางและเป็นประโยชน์ต่อวงการการศึกษา ผู้วิจัยจึงนำประเด็นข้อสอบที่มีการตรวจให้คะแนนแบบผสมผสาน (mixed format test) มาพัฒนาองค์ความรู้เพิ่มเติมให้เข้ากับบริบทของประเทศไทยมากยิ่งขึ้น

ตอนที่ 2 มโนทัศน์การทำหน้าที่ต่างกันเป็นขั้น

ในตอนที่สองผู้วิจัยแบ่งการนำเสนอออกเป็น 6 ส่วนได้แก่ ที่มาของการทำหน้าที่ต่างกันเป็นขั้น การอธิบายขั้น (step) ตามแนวคิดการทำหน้าที่ต่างกันเป็นขั้น รูปแบบของผลการทำหน้าที่ต่างกันเป็นขั้น เกณฑ์ในการกำหนดรูปแบบการทำหน้าที่ต่างกันเป็นขั้น วิธีการเชิงสถิติสำหรับการตรวจสอบการทำหน้าที่ต่างกันเป็นขั้น การใช้ผลการทำหน้าที่ต่างกันเป็นขั้นเพื่อทดสอบการทำหน้าที่ต่างกันของข้อสอบ และงานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันเป็นขั้น โดยมีรายละเอียดดังต่อไปนี้

2.1 ที่มาของการทำหน้าที่ต่างกันเป็นขั้น (differential step functioning: DSF)

แนวคิดเดิมของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในข้อสอบที่มีการตรวจให้คะแนนพหุวิภาคเป็นเพียงการตรวจสอบความแตกต่างระหว่างกลุ่มในระดับข้อสอบ แต่ไม่ได้ระบุขั้นตอนการให้คะแนนที่อาจจะเกิดการทำหน้าที่ต่างกันของข้อสอบขึ้น ซึ่งการระบุการทำหน้าที่ต่างกันเป็นขั้นสามารถให้ข้อมูลสารสนเทศที่จำเป็นต่อการเกิดข้อสอบทำหน้าที่ต่างกันในแต่ละขั้นของการตรวจให้คะแนน จากเหตุผลดังกล่าว จึงมีนักวิชาการที่พยายามศึกษาและเพิ่มสารสนเทศการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนพหุวิภาค โดยนำเสนอการศึกษาการทำหน้าที่ต่างกันเป็นขั้น (differential step functioning: DSF) ซึ่งเป็นวิธีที่ใช้ในการตรวจสอบความแตกต่างระหว่างกลุ่ม (between-group differences) ในแต่ละขั้น (step) ของข้อสอบที่มีการตรวจให้คะแนนพหุวิภาค (Penfield, Gattamorta, & Childs, 2009; Penfield, Alvarez, & Lee, 2008)

กรอบการทำหน้าที่ต่างกันเป็นขั้นซึ่งมีการให้คะแนนครอบคลุมการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบที่มีประสิทธิภาพมากกว่าการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบหลายประการด้วยกัน ประการแรก ผลการทำหน้าที่ต่างกันเป็นขั้น (the DSF effects) มีอำนาจในการตรวจสอบมากกว่าการทดสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อขนาด (magnitude) และ/หรือเครื่องหมาย (sign) ของผลการทำหน้าที่ต่างกันเป็นขั้นสามารถแปรเปลี่ยนข้ามขั้นได้ ประการที่สอง ผลการทำหน้าที่ต่างกันเป็นขั้นสามารถแยกวิเคราะห์เป็นรายข้อที่สามารถให้ข้อปรับปรุงในเนื้อหาหรือข้อเสนอแนะปัจจัยในการเกิดการทำหน้าที่ต่างกันของข้อสอบ เนื่องจากวิธีการทำหน้าที่ต่างกันเป็นขั้นเป็นการระบุความแตกต่างระหว่างกลุ่มในข้อสอบที่การตรวจให้คะแนนพหุวิภาค (Penfield, 2007a, 2007b; Penfield, Gattamorta, & Childs, 2009; Penfield, Alvarez, & Lee, 2008)

2.2 การอธิบายขั้น (step) ตามแนวคิดการทำหน้าที่ต่างกันเป็นขั้น

ตามโมเดลทฤษฎีการตอบสนองข้อสอบอนุกรมของโมเดลอย่างง่ายเชิงสัมพัทธ์ (a series of relatively simple models) ถูกเรียกว่า การทำหน้าที่เป็นขั้น (step functions) ใช้อธิบายโอกาสที่ผู้เข้าสอบเฉพาะระดับขั้นความสามารถ หรือ การได้เปรียบ (advance) จากคะแนนในลำดับที่หนึ่งไปจนถึงคะแนนในระดับที่สูงกว่าในข้อสอบที่มีการให้คะแนนพหุวิภาค (Penfield, Gattamorta, & Childs, 2009)

ในแต่ละขั้นสามารถอธิบายได้ตามทฤษฎีตอบสนองข้อสอบโลจิสติก 2 พารามิเตอร์ (two-parameter logistic IRT) ซึ่งภายในโมเดลประกอบไปด้วยพารามิเตอร์ความยาก (b_j) เป็นพารามิเตอร์ของแต่ละขั้น และพารามิเตอร์อำนาจจำแนก (a) เป็นพารามิเตอร์คงที่ในทุกๆ ขั้น และฟังก์ชันความสามารถ (θ) ซึ่งรูปแบบการทำหน้าที่เป็นขั้นที่ขั้นที่ J อธิบายได้ตามสมการที่ 2.3

$$P(Y \geq j \mid \theta) = \frac{\exp[a(\theta - b_j)]}{1 + \exp[a(\theta - b_j)]} \quad \text{.....(2.3)}$$

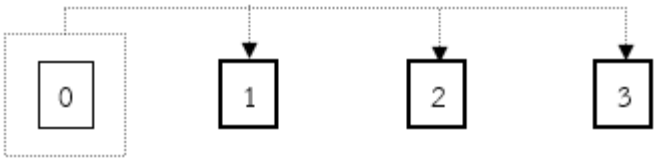
โดยที่

a หมายถึง พารามิเตอร์ค่าอำนาจจำแนก

b_j หมายถึง พารามิเตอร์ความยากในแต่ละขั้น

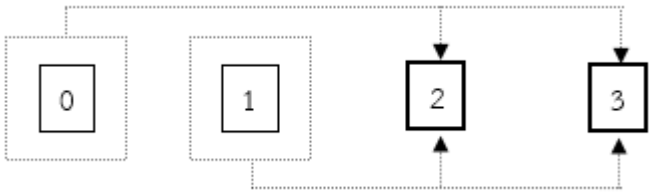
นอกจากนี้ในข้อสอบที่มีการตรวจให้คะแนนพหุวิภาค (polytomous items) ที่มีลำดับการให้คะแนน r ลำดับ จะมีการทำหน้าที่เป็นขั้นอยู่ r-1 ขั้น เช่น ในข้อสอบที่มีการตรวจให้คะแนน 4 ระดับสามารถแสดงการทำหน้าที่เป็นขั้นได้ 3 ขั้น ทั้งนี้ลักษณะการเกิดขั้นในข้อสอบที่มีการตรวจให้คะแนน 4 ระดับ เริ่มจากคะแนน 0, 1, 2, 3, และ 4 สามารถแยกขั้นการเกิดในแต่ละขั้นได้ (Agresti, 1990 cited in Penfield, Alvarez, & Lee, 2008) โดยที่การทำหน้าที่ในขั้นที่ 1 เป็นฟังก์ชันความสามารถที่เกิดจากความน่าจะเป็นของก้าวผ่านจากคะแนนที่ 0 ไปยังคะแนน 1, 2 หรือ 3 แสดงได้ตามรูปภาพที่ 2

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



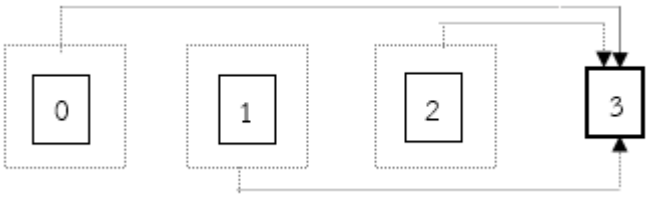
รูปภาพที่ 2 แสดงการทำหน้าที่เป็นขั้นในขั้นที่ 1

ส่วนการทำหน้าที่ในขั้นที่ 2 เป็นฟังก์ชันความสามารถที่เกิดจากความน่าจะเป็นของก้าวผ่านจากคะแนนที่ 0 หรือ 1 ไปยังคะแนน 2 หรือ 3 แสดงได้ตามรูปภาพที่ 3



รูปภาพที่ 3 แสดงการทำหน้าที่เป็นขั้นในขั้นที่ 2

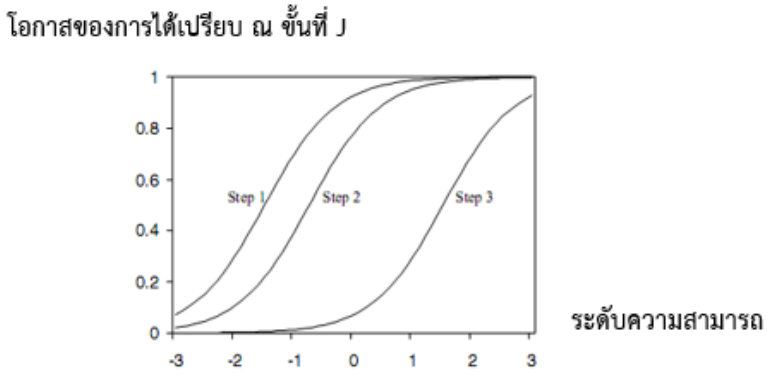
และการทำหน้าที่ในขั้นที่ 3 เป็นฟังก์ชันความสามารถที่เกิดจากความน่าจะเป็นของการก้าวผ่านคะแนนที่ 0, 1 หรือ 2 ไปยังคะแนน 3 แสดงได้ตามรูปภาพที่ 4



จุฬาลงกรณ์มหาวิทยาลัย

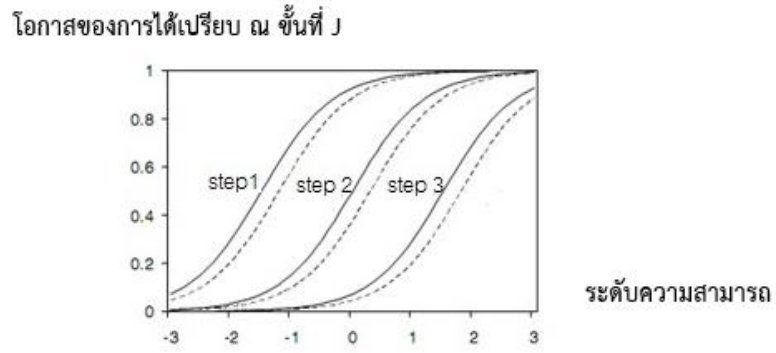
รูปภาพที่ 4 แสดงการทำหน้าที่เป็นขั้นในขั้นที่ 3

ส่วนขนาดของผลการทำหน้าที่ต่างกันเป็นขั้นแบบคงที่ทุกๆ ขั้น แสดงได้ตามรูปภาพที่ 5



รูปภาพที่ 5 แสดงการทำหน้าที่เป็นขั้นในข้อสอบที่มีการตรวจให้คะแนน 4 ค่า มาจาก Penfield, Gattamorta, และ Childs (2009)

ในกรณีที่พารามิเตอร์ความยากมีความแตกต่างระหว่างกลุ่มอ้างอิง (Reference: R) และกลุ่มเปรียบเทียบ (Focal: F) ลักษณะการเกิดการทำหน้าที่ต่างกันเป็นขั้นของทั้งสองกลุ่มมีลักษณะเหมือนกัน b_{jR} จะเท่ากับ b_{jF} ในแต่ละขั้น แต่ถ้าพารามิเตอร์ความยากของกลุ่มเปรียบเทียบมากกว่ากลุ่มอ้างอิง เช่น พารามิเตอร์ความยากของกลุ่มเปรียบเทียบมากกว่ากลุ่มอ้างอิงอยู่ .3 หน่วย จะได้ว่า $b_{jF} = b_{jR} + .3$ หมายความว่า ในระดับความสามารถเดียวกันโอกาสของ การได้เปรียบของแต่ละขั้นไม่ได้เท่ากันและในกลุ่มเปรียบเทียบมีขนาดมากกว่ากลุ่มอ้างอิง .3 หน่วย ดังนั้นขนาดของผลการทำหน้าที่ต่างกันเป็นขั้นจึงมีขนาดเท่ากับ .3 แสดงได้ตามรูปภาพที่ 6



รูปภาพที่ 6 แสดงการทำหน้าที่เป็นขั้นในข้อสอบที่มีการตรวจให้คะแนน 4 ค่า ในกลุ่มอ้างอิง (เส้นทึบ) และกลุ่มเปรียบเทียบ(เส้นประ) มาจาก Penfield, Gattamorta, และ Childs, 2009)

2.3 รูปแบบของผลการทำหน้าที่ต่างกันเป็นขั้น

การทำหน้าที่ต่างกันเป็นขั้นและการทำหน้าที่ต่างกันของข้อสอบในข้อสอบที่มีการตรวจให้คะแนนพหุวิภาคมีความเชื่อมโยงกัน โดยแนวคิดการทำหน้าที่ต่างกันเป็นขั้นเป็นการรวมผลการทำหน้าที่ต่างกันของข้อสอบในทุกๆ ขั้น นอกจากนี้รูปแบบการทำหน้าที่ต่างกันเป็นขั้นที่สามารถเชื่อมโยงผลการทำหน้าที่ต่างกันของข้อสอบสามารถแบ่งได้เป็น 2 มิติ (Penfield, Alvarez, & Lee, 2008) ได้แก่ มิติที่หนึ่งอธิบายตามการจัดกลุ่มภายในรูปแบบการทำหน้าที่ต่างกันเป็นขั้น และมิติที่สองผลการทำหน้าที่ต่างกันเป็นขั้น ซึ่งทั้งสองมิติสามารถช่วยระบุสาเหตุการทำหน้าที่ต่างของข้อสอบ (DIF) และช่วยในการตัดสินใจว่าจะคงข้อสอบหรือกำจัดข้อสอบออกทิ้งไป รายละเอียดดังต่อไปนี้

มิติที่หนึ่งการจัดกลุ่มภายในรูปแบบการทำหน้าที่ต่างกันเป็นขั้น มิติการจัดกลุ่มภายในรูปแบบการทำหน้าที่ต่างกันเป็นขั้นแบ่งได้เป็น 2 ลักษณะ คือ ลักษณะกระจาย (pervasive) และลักษณะไม่กระจาย (nonpervasive) รายละเอียดตามตารางที่ 3

ตารางที่ 3 การจัดกลุ่มภายในรูปแบบการทำหน้าที่ต่างกันเป็นขั้น

การจัดกลุ่ม	ลักษณะการเกิด DSF	`ความเชื่อมโยงกับ DIF
การทำหน้าที่ต่างกันเป็นขั้น ลักษณะกระจาย (pervasive DSF)	ทุกๆ ขั้นแสดงผลของการทำหน้าที่ต่างกันเป็นขั้น และผลการของการทำหน้าที่ต่างกันเป็นขั้นกระจายทุกๆ ระดับคะแนน	สาเหตุของการทำหน้าที่ต่างกันของข้อสอบมีอิทธิพลในระดับข้อสอบ (the item level)
การทำหน้าที่ต่างกันเป็นขั้น ลักษณะไม่กระจาย (nonpervasive DSF)	ขั้นเพียงหนึ่งหรือสองขั้นแสดงผลของการทำหน้าที่ต่างกันเป็นขั้น	การทำหน้าที่ต่างกันของข้อสอบอาจจะถูกแยกเป็นหนึ่งหรือสองขั้น

มิติที่สองผลการทำหน้าที่ต่างกันเป็นขั้น มิติผลการทำหน้าที่ต่างกันเป็นขั้นแบ่งได้เป็น 3 ลักษณะ ได้แก่ ลักษณะคงที่ (constant) ลักษณะลู่เข้า (convergent) และลักษณะลู่ออก (divergent) รายละเอียดตามตารางที่ 4

ตารางที่ 4 ผลการทำหน้าที่ต่างกันเป็นขั้น

ผลการเกิด DSF	ลักษณะการเกิด DSF
ผลการทำหน้าที่ต่างกันเป็นขั้นลักษณะคงที่ (constant DSF)	การทำหน้าที่ต่างกันเป็นขั้นที่เครื่องหมาย (sign) และขนาด (magnitude) เดียวกัน
ผลการทำหน้าที่ต่างกันเป็นขั้นลักษณะลู่เข้า (convergent DSF)	ผลของการทำหน้าที่ต่างกันเป็นขั้นมีเครื่องหมายเดียวกัน (sign) แต่มีขนาด (magnitude) ที่แตกต่างกัน
ผลการทำหน้าที่ต่างกันเป็นขั้นลักษณะลู่ออก (divergent DSF)	ลักษณะเครื่องหมายมีลักษณะที่ตรงกันข้าม หรือการเปลี่ยนเครื่องหมายระหว่างกลุ่มข้ามขั้น

จากมิติการจัดกลุ่มภายในรูปแบบการทำหน้าที่ต่างกันเป็นขั้น และมิติผลการทำหน้าที่ต่างกันเป็นขั้น สามารถรวมมิติทั้งสองเพื่อจัดหมวดหมู่การทำหน้าที่ต่างกันเป็นขั้น (a taxonomy of DSF) แบ่งออกเป็น 2 มิติ คือ มิติที่หนึ่งตำแหน่งของการเกิดผลการทำหน้าที่ต่างกันเป็นขั้น และมิติที่สองความสอดคล้องของอิทธิพลเชิงสาเหตุที่แสดงผลการทำหน้าที่ต่างกันเป็นขั้น (Penfield, Alvarez, & Lee, 2008) รายละเอียดตามตารางที่ 5

ตารางที่ 5 ลักษณะการทำหน้าที่ต่างกันเป็นขั้นตามหมวดหมู่ (a taxonomy of DSF)

ความสอดคล้อง (ความสอดคล้องของสาเหตุการเกิด DIF)	**การกระจาย (ตำแหน่งของสาเหตุการเกิด DIF)**	
	กระจาย (pervasive)	**ไม่กระจาย (Non-pervasive)**
ลักษณะคงที่ (constant)	ทุกขั้นแสดงการเกิดผลการทำหน้าที่ต่างกันเป็นขั้นโดยมีขนาดและเครื่องหมายเท่ากัน	มีเพียงหนึ่งหรือสองขั้นแสดงการเกิดผลการทำหน้าที่ต่างกันเป็นขั้นโดยมีขนาดและเครื่องหมายเท่ากัน
ลักษณะลู่เข้า (convergent)	ทุกขั้นแสดงการเกิดผลการทำหน้าที่ต่างกันเป็นขั้นมีเครื่องหมายเดียวกันแต่ขนาดต่างกัน	มีเพียงสองขั้นแสดงผลการเกิดการทำหน้าที่ต่างกันเป็นขั้นที่มีเครื่องหมายเดียวกัน แต่ขนาดแตกต่างกัน
ลักษณะลู่ออก (Divergent)	ทุกขั้นแสดงการเกิดผลการทำหน้าที่ต่างกันเป็นขั้นแต่เครื่องหมายแปรเปลี่ยนข้ามขั้น	มีเพียงสองขั้นแสดงผลการทำหน้าที่ต่างกันเป็นขั้นและเครื่องหมายแปรเปลี่ยนข้ามขั้น

จากตาราง 2.5 กล่าวโดยสรุปคือ **มิติที่หนึ่ง** แสดงความแตกต่างระหว่าง ลักษณะกระจาย (pervasive) และ ไม่กระจาย (non-pervasive) โดยที่ **การทำหน้าที่ต่างกันเป็นขั้นลักษณะกระจาย (pervasive DSF)** สามารถเกิดการทำหน้าที่ต่างกันของข้อสอบในระดับของข้อสอบมากกว่าในระดับขั้น **ส่วนการทำหน้าที่ต่างกันเป็นขั้นลักษณะไม่กระจาย (Non-**

pervasive DSF) สัมพันธ์กับบางสถานการณ์ไม่ใช่เกิดในทุกๆ ขั้น แสดงให้เห็นว่า สาเหตุการเกิดการทำหน้าที่ต่างกันของข้อสอบอาจจะเกิดเพียงหนึ่งขั้น สองขั้น หรือสามขั้นก็ได้

มิติที่สอง แสดงความสอดคล้องของผลการทำหน้าที่ต่างกันเป็นขั้นในลักษณะคงที่ (constant) ลู่เข้า (convergent) และลู่ออก (divergent) โดยลักษณะคงที่ (constant) แสดงให้เห็นสถานการณ์ที่ทุกๆ ขั้น แสดงผลการทำหน้าที่ต่างกันเป็นขั้นที่สัมพันธ์กันทั้งขนาด (magnitude) และเครื่องหมาย (sign) สถานการณ์การกระจายแบบคงที่ (constant pervasive DSF) ต้องมีหลักฐานที่ชัดเจนว่าผลการทำหน้าต่างกันของข้อสอบเกิดมาจากคุณสมบัติระดับข้อสอบเท่านั้น ส่วนสถานการณ์การไม่กระจายแบบคงที่ (constant non-pervasive DSF) สาเหตุการเกิดการทำหน้าที่ต่างกันของข้อสอบไม่จำเป็นที่คุณสมบัติระดับข้อสอบ (item-level property) แต่เป็นคุณสมบัติที่ประกอบด้วยทุกๆ ขั้น ที่เกิดผลการเกิดการทำหน้าที่ต่างกันเป็นขั้น (DSF)

การทำหน้าที่ต่างกันเป็นขั้นแบบลู่เข้า (convergent DSF) เกิดขึ้นเมื่อทุกๆ ขั้น เกิดผลการทำหน้าที่ต่างกันเป็นขั้นที่มีเครื่องหมายเดียวกันแต่มีขนาดไม่เท่ากัน โดยลักษณะการเกิดดังกล่าวพบว่า คุณสมบัติสาเหตุการเกิดการทำหน้าที่ต่างกันของข้อสอบเข้าข้างกลุ่มเดียวกันในทุกๆ ขั้น เป็นไปได้ว่าสาเหตุของการทำหน้าที่ต่างกันของข้อสอบอยู่ในระดับข้อสอบและผลกระทบของสาเหตุขึ้นอยู่กับคุณสมบัติของระดับคะแนนแต่ละระดับ หรือเป็นไปได้ว่ามีมากกว่าหนึ่งคุณสมบัติที่แตกต่างกันในระดับคะแนน

การทำหน้าที่ต่างกันเป็นขั้นลักษณะลู่ออก (divergent DSF) เกิดขึ้นเมื่อทุกๆ ขั้นมีเครื่องหมายที่แตกต่างกัน ลักษณะการเกิดดังกล่าวชี้ให้เห็นว่าเกิดการได้เปรียบข้ามระหว่างกลุ่มในแต่ละขั้น โดยการเกิดประเภทนี้ทำให้ทราบสาเหตุของการทำหน้าที่ต่างกันเป็นขั้นเป็นที่เป็นลักษณะเฉพาะของแต่ละขั้น

CHULALONGKORN UNIVERSITY

2.4 เกณฑ์ในการกำหนดรูปแบบการทำหน้าที่ต่างกันเป็นขั้น

การระบุรูปแบบการทำหน้าที่ต่างกันเป็นขั้น ได้แก่ แบบกระจาย (pervasive) ไม่กระจาย (non-pervasive) แบบคงที่ (constant) แบบลู่เข้า (convergent) และแบบลู่ออก (divergent) ขึ้นอยู่กับสัญลักษณ์ (sign) และขนาด (magnitude) นอกจากนี้วิธีวิทยาการในการพัฒนาการประมาณค่าการทำหน้าที่ต่างกันเป็นขั้นประกอบไปด้วย ทฤษฎีการตอบสนองข้อสอบ (IRT) วิธีถดถอยโลจิสติก (logistic regression) และวิธีอัตราส่วนเลขออด (odds ratio approaches) โดยเกณฑ์ที่ Penfield นำเสนอในการกำหนดรูปแบบการทำหน้าที่ต่างกันเป็นขั้นใช้แนวคิดวิธีอัตราส่วนเลขออด ทั้งนี้ในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบในรูปแบบข้อสอบประเภทตรวจให้คะแนนทวิวิภาค (dichotomous items) และตรวจให้คะแนนพหุวิภาค (polytomous items)

สามารถคำนวณได้จากอัตราส่วนลอกเลขออดแต้มต่อร่วม (the common log-odds ratio) (Penfield, Alvarez, & Lee, 2008) ตามสมการที่ 2.4

$$\hat{\lambda}_j = \ln\left[\frac{\sum_{k=1}^{m}\frac{A_{jk}D_{jk}}{T_k}}{\sum_{k=1}^{m}\frac{B_{jk}C_{jk}}{T_k}}\right] \qquad \text{.....(2.4)}$$

โดยที่

A_{jk} แทน จำนวนสมาชิกของกลุ่มอ้างอิง ณ ความสามารถ ระดับที่ k โดยเป็นคนที่ประสบผลสำเร็จที่ระดับ j

B_{jk} แทน จำนวนสมาชิกของกลุ่มอ้างอิง ณ ความสามารถ ระดับที่ k โดยเป็นคนที่ไม่ประสบผลสำเร็จที่ระดับ j

C_{jk} แทน จำนวนสมาชิกของกลุ่มเปรียบเทียบ ณ ความสามารถ ระดับที่ k โดยเป็นคนที่ประสบผลสำเร็จที่ระดับ j

D_{jk} แทน จำนวนสมาชิกของกลุ่มเปรียบเทียบ ณ ความสามารถ ระดับที่ k โดยเป็นคนที่ไม่ประสบผลสำเร็จที่ระดับ j

T_k แทน จำนวนสมาชิกทั้งกลุ่มอ้างอิงและเปรียบเทียบทั้งหมด ณ ความสามารถระดับ ที่ k

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

โดยที่ ลอกกาลิทึม ของ $\hat{\alpha}_j$ แทนด้วย $\hat{\lambda}_j$ นอกจากนี้ $\hat{\lambda}_j$ ยังสามารถระบุการทำหน้าที่ต่างกันเป็นขั้น(DSF) ได้ ซึ่ง $\hat{\lambda}_j$ มีค่าเป็นบวก หมายถึง การทำหน้าที่ต่างกันเป็นขั้นเข้าข้างกลุ่มอ้างอิง ส่วน $\hat{\lambda}_j$ มีค่าเป็นลบ หมายถึง การทำหน้าที่ต่างกันเป็นขั้นเข้าข้างกลุ่มเปรียบเทียบ ถ้า $\hat{\lambda}_j$ มีค่าเป็นศูนย์ หมายถึง ไม่เกิดการทำหน้าที่ต่างกันเป็นขั้น ทั้งนี้ตัวประมาณค่าความคลาดเคลื่อนมาตรฐานของ $\hat{\lambda}_j$ สามารถอธิบายได้ตามสมการที่ 2.5

$$SE(\hat{\lambda}_j) \sqrt{\frac{\sum_{k=1}^{m} T_k^{-2}(A_{jk}D_{jk} + \hat{\alpha}_j B_{jk}C_{jk})(A_{jk} + D_{jk} + \hat{\alpha}_j B_{jk} + \hat{\alpha}_j C_{jk}}{2\left(\sum_{k=1}^{m}\frac{A_{jk}D_{jk}}{T_{jk}}\right)}} \qquad \text{.....(2.5)}$$

ทั้งนี้ตัวประมาณค่าความคลาดเคลื่อนมาตรฐานของ $\hat{\lambda}_j$ มีลักษณะเดียวกับ อัตราส่วนลอกเลขออดร่วมของแมนเทลและแฮนด์เซล (the mantel-haenszel common log odds ratio: $\hat{\lambda}_{MH}$) (Camilli & Shepard, 1994; Penfield & Camilli, 2007 cited in Gattamorta & Penfield, 2012) และสถิติที่ใช้ทดสอบ $\hat{\lambda}_j$ อธิบายได้ตามสมการที่ 2.6

$$Z(\hat{\lambda}_j) = \frac{\hat{\lambda}_j}{SE(\hat{\lambda}_j)} \quad \text{.....(2.6)}$$

ตามวิธีการแบ่งหมวดหมู่ของ ETS สามารถประยุกต์ใช้หลักการอัตราส่วนลอกเลขออดร่วมขั้นเป็นระดับ (the step-level common log-odds ratio) สำหรับขั้นที่ j ใดๆ ($\hat{\lambda}_j$) โดยการเกิดการทำหน้าที่ต่างกันเป็นขั้นสามารถแบ่งได้เป็น ขนาดเล็ก ขนาดกลาง และขนาดใหญ่ โดยพิจารณาลักษณะการเกิดทั้งแบบลู่เข้า (convergent) และ แบบลู่ออก (divergent) สามารถแบ่งหมวดหมู่การแบ่งผลการทำหน้าที่ต่างกันเป็นขั้น รายละเอียดตามตารางที่ 6

ตารางที่ 6 สัญลักษณ์และเกณฑ์การแปลความหมายของ $\hat{\lambda}_j$

สัญลักษณ์	คำอธิบาย	ช่วงพิจารณา
S	ผลการทำหน้าที่ต่างกันเป็นขั้นขนาดเล็ก	$\lvert\hat{\lambda}_j\rvert < 0.43$
M-	ผลการทำหน้าที่ต่างกันเป็นขั้นขนาดกลางทางด้านลบ	$-0.64 < \hat{\lambda}_j < -0.43$
M+	ผลการทำหน้าที่ต่างกันเป็นขั้นขนาดกลางทางด้านบวก	$0.43 < \hat{\lambda}_j < 0.64$
L-	ผลการทำหน้าที่ต่างกันเป็นขั้นขนาดใหญ่ทางด้านลบ	$\hat{\lambda}_j \leq -0.64$
L+	ผลการทำหน้าที่ต่างกันเป็นขั้นขนาดใหญ่ทางด้านบวก	$\hat{\lambda}_j \geq 0.64$

เมื่อนำกลุ่มที่แบ่งทั้ง 5 กลุ่ม แยกอธิบายตามหมวดหมู่การทำหน้าที่ต่างกันเป็นขั้น สามารถอธิบายรูปแบบการเกิดการทำหน้าที่ต่างกันเป็นขั้นโดยใช้พื้นฐานการจัดกลุ่มตามผลการทำหน้าที่ต่างกันเป็นขั้น รายละเอียดตามตารางที่ 7

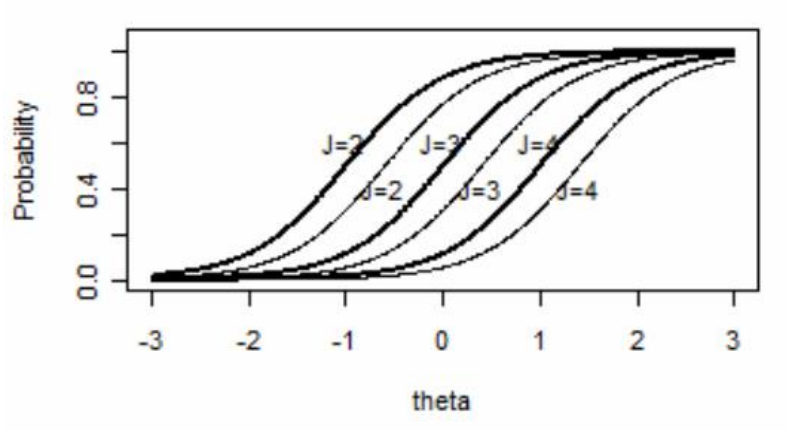
ตารางที่ 7 รูปแบบการทำหน้าที่ต่างกันตามการจัดกลุ่มตามผลการทำหน้าที่ต่างกันเป็นขั้น (a DSF effect categorization scheme)

รูปแบบ DSF	คำอธิบาย	ตัวอย่างข้อสอบ 3 ที่มีขั้น		
		ขั้นที่ 1	ขั้นที่ 2	ขั้นที่ 3
no DSF	ทุกขั้นมีขนาด S	S	S	S
pervasive constant	ทุกขั้น ทั้งหมดเป็น M- หรือ ทั้งหมดเป็น M+ หรือ ทั้งหมดเป็น L- หรือ ทั้งหมดเป็น L+	M±, L±	M±, L±	M±, L±
pervasive convergent	ทุกขั้น ทั้งหมดเป็น M- และ L- ***หรือ*** ทุกขั้นทั้งหมดเป็น M+ และ L+	M-	M-	L-
pervasive divergent	ไม่มีขั้นใดเป็น S ***และ*** อย่างน้อยหนึ่งขั้นเป็น M- หรือ L- ***และ*** อย่างน้อยหนึ่งขั้นเป็น M+ หรือ L+	M+	L+	L-
non-pervasive constant	อย่างน้อยหนึ่งขั้นเป็น S ***และ*** ทุกๆ ขั้นทั้งหมดเป็น M- หรือทุกๆ ขั้นทั้งหมดเป็น M+ หรือทุกๆ ขั้นทั้งหมดเป็น L- หรือทุกๆ ขั้นทั้งหมดเป็น L+	S	L-	L-
non-pervasive convergent	อย่างน้อยหนึ่งขั้นเป็น S ***และ*** ทุกๆ ขั้นเป็นการผสมระหว่าง M- และ L- หรือ เป็นการผสมระหว่าง M+ และ L+	S	M+	L+
non-pervasive divergent	อย่างน้อยหนึ่งขั้นเป็น S ***และ*** อย่างหนึ่งขั้นเป็น M- หรือ L- ***และ*** อย่างน้อยหนึ่งขั้นเป็น M+ หรือ L+	S	L+	L-

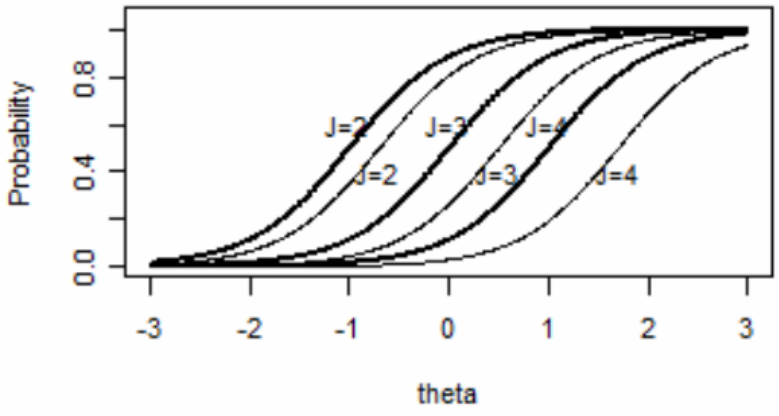
หมายเหตุ

S หมายถึง small DSF effect , M- หมายถึง medium negative DSF effect , M+ หมายถึง medium positive DSF effect , L- หมายถึง large negative DSF effect และ L+ หมายถึง large positive DSF effect มาจาก Penfield, Alvarez, & Lee (2008)

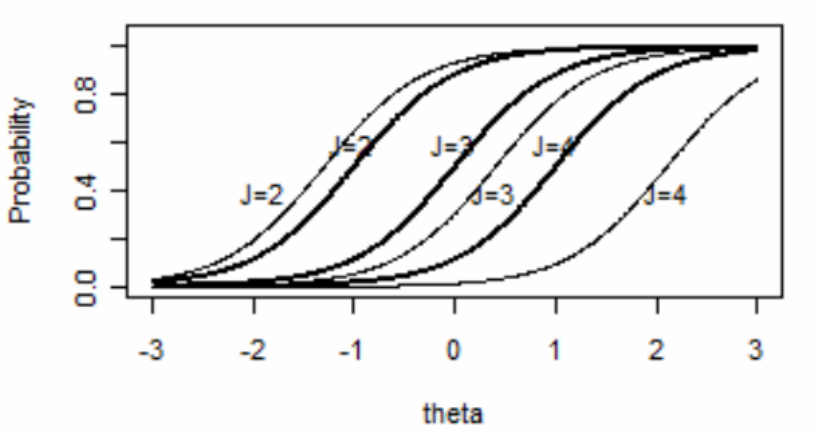
ลักษณะการเกิดการทำหน้าที่ต่างกันเป็นขั้นบนพื้นฐานการจัดกลุ่มตามผลการทำหน้าที่ต่างกันเป็นขั้นสำหรับข้อสอบที่มีการตรวจให้คะแนนพหุวิภาคที่มี 4 ค่า สามารถแสดงการเกิดตามรูปภาพที่ 7 ดังนี้



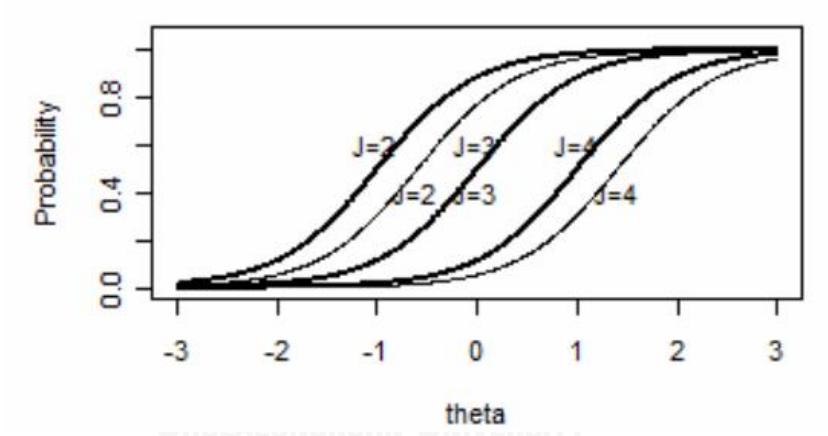
การเกิดแบบ pervasive constant DIF



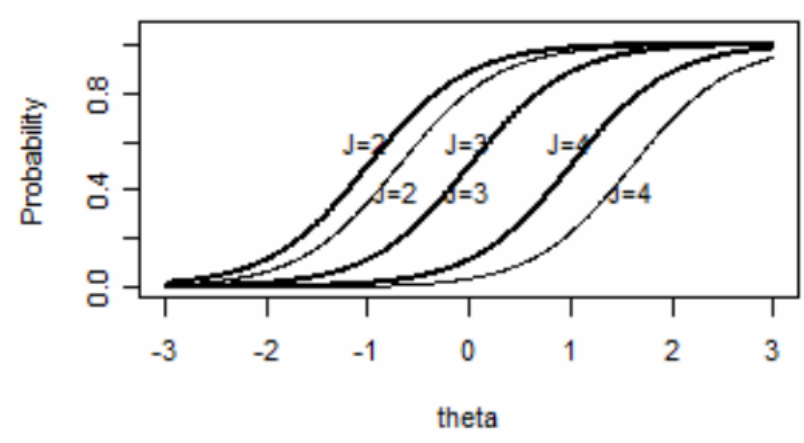
การเกิดแบบ pervasive convergent DIF



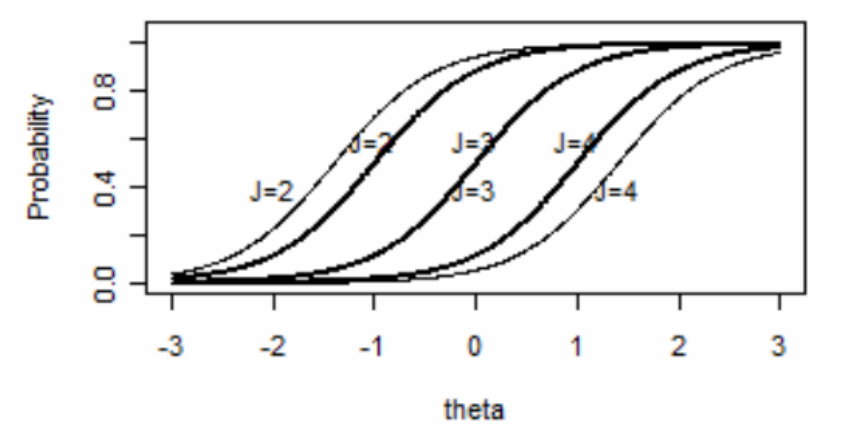
การเกิดแบบ pervasive divergent DIF



การเกิดแบบ non-pervasive constant DIF



การเกิดแบบ non-pervasive convergent DIF



การเกิดแบบ non-pervasive convergent DIF

รูปภาพที่ 7 แสดงผลการทำหน้าที่ต่างกันเป็นขั้นในข้อสอบที่มีการตรวจให้คะแนนพหุวิภาค 4 ค่า พร้อมการเกิด 6 ลักษณะ มาจาก Wood (2011)

2.5 วิธีการเชิงสถิติสำหรับการตรวจสอบการทำหน้าที่ต่างกันเป็นขั้น

วิธีการเชิงสถิติที่ใช้ในการตรวจสอบการทำหน้าที่ต่างกันเป็นขั้น ประกอบด้วย วิธีทฤษฎีการตอบสนองข้อสอบ (IRT approach) วิธีอัตราส่วนเลขออด (odds ratio approach) และวิธีถดถอยโลจิสติก (logistic regression approach) ซึ่งทั้งสามวิธีเป็นแนวคิดแบบ cumulative approach และใช้ในการตรวจสอบการทำหน้าที่ต่างกันเป็นขั้นในแต่ละขั้นของข้อสอบ รายละเอียดดังต่อไปนี้

2.5.1 วิธีทฤษฎีตอบสนองข้อสอบ (IRT approach) สำหรับการทำหน้าที่ต่างกันเป็นขั้นอยู่บนพื้นฐานการทดสอบความแตกต่างระหว่างกลุ่ม (between group differences) ผลการทำหน้าที่ต่างกันเป็นขั้นที่ขั้นที่ J สามารถให้นิยามเป็นความแตกต่างระหว่างกลุ่มสำหรับพารามิเตอร์ b_j ตามสมการที่ 2.7

$$\Delta (b_j) = b_{jF} - b_{jR} \quad \text{.....(2.7)}$$

โดยที่

$\Delta (b_j) = 0$ หมายถึง การไม่เกิดการทำหน้าที่ต่างกันของขั้น (no DSF) สำหรับขั้นที่ J

$\Delta (b_j) > 0$ หมายถึง การได้เปรียบเชิงสัมพัทธ์สำหรับกลุ่มอ้างอิงขั้นที่ J

$\Delta (b_j) < 0$ หมายถึง การได้เปรียบเชิงสัมพัทธ์สำหรับกลุ่มเปรียบเทียบขั้นที่ J

การทดสอบสมมติฐานว่างของการไม่เกิดการทำหน้าที่ต่างกันของขั้น (no DSF) ตามทฤษฎีตอบสนองข้อสอบสามารถใช้ได้ 2 วิธี คือ วิธีแรก $\Delta (b_j)$ หารด้วย ความคลาดเคลื่อนมาตรฐานเพื่อให้อยู่ในรูปมาตรฐานทั่วไป วิธีที่สอง ใช้วิธี likelihood ratio test ทั้งนี้ $\Delta (b_j)$ แสดงถึงพื้นที่คิดเครื่องหมายระหว่างการทำหน้าที่เป็นขั้นที่ขั้นที่ J ของทั้งกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ โดยที่ค่า b_j ของทั้งสองกลุ่มเป็นค่าที่ได้จากการประมาณค่าผลการทำหน้าที่ต่างกันเป็นขั้นสำหรับขั้นที่ J (Cohen, Kim, & Baker, 1993 cited in Penfield, Gattamorta, & Childs, 2009)

2.5.2 วิธีอัตราส่วนเลขออด (odds ratio approach) เป็นวิธีที่ใช้ตรวจสอบผลการทำหน้าที่ต่างกันของขั้นโดยการเปรียบเทียบเลขออดที่ได้เปรียบอย่างสมบูรณ์ ณ ขั้น ที่ J ทั้งกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่ระดับความสามารถเดียวกัน (Penfield, 2007a; Penfield, Gattamorta, & Childs, 2009) ความเท่าเทียมกันระหว่างกลุ่ม ของเลขออดการได้เปรียบในขั้นที่ J สามารถตรวจสอบโดยพิจารณาอัตราส่วนเลขออดของความสำเร็จของกลุ่มอ้างอิงที่มากกว่ากลุ่มเปรียบเทียบ ค่าของอัตราส่วนเลขออดในทุกๆ ระดับความสามารถ เรียกว่า อัตราส่วนเลขออดร่วม (a common odds ratio) และ ลอกการิทึ่ม ของอัตราส่วนเลขออดร่วม สำหรับขั้นที่ J แทนด้วย λ_j การแปลผลมีรายละเอียดดังนี้

$\lambda_j = 0$ หมายถึง การไม่เกิดการทำหน้าที่ต่างกันของขั้น สำหรับขั้นที่ J

$\lambda_j > 0$ หมายถึง การทำหน้าที่ต่างกันของขั้นโดยเข้าข้างกลุ่มอ้างอิงขั้นที่ J

$\lambda_j < 0$ หมายถึง การทำหน้าที่ต่างกันของขั้นโดยเข้าข้างกลุ่มเปรียบเทียบขั้นที่ J

การทดสอบสมมติฐานการไม่เกิดการทำหน้าที่ต่างกันของขั้น ที่ขั้นที่ J (no DSF) โดยใช้วิธีเลขออดสามารถทดสอบตามสมการที่ 2.8

$$z(\hat{\lambda}j) = \frac{\hat{\lambda}j}{SE(\hat{\lambda}j)} \quad \text{.....(2.8)}$$

2.5.3 วิธีถดถอยโลจิสติก (logistic regression approach) เป็นวิธีที่ใช้สำหรับตรวจสอบการทำหน้าที่ต่างกันเป็นขั้นอยู่บนพื้นฐานโมเดลฟังก์ชันเป็นขั้นที่โอกาสการเพิ่มที่ขั้น J ใดๆ (the probability of successfully advancing at the jth) เป็นฟังก์ชันของตัวแปรคะแนนแบบสอบที่ถูกสังเกต (X) และสมาชิกกลุ่ม (G) (French & Miller, 1996 cited in Penfield, Gattamorta, & Childs, 2009) อธิบายได้ตามสมการที่ 2.9

$$P(Y \geq j|X) = \frac{\exp(\beta_{j0} + \beta j_{1X} + \beta_{j2} X)}{1 + \exp(\beta_{j0} + \beta j_{1X} + \beta_{j2} X)} \quad \text{.....(2.9)}$$

โดยที่

G หมายถึง ตัวแปรกลุ่ม โดยที่ G=1 เป็นกลุ่มอ้างอิง และ G=0 เป็นกลุ่มเปรียบเทียบ

β_{j2} หมายถึง ผลการทำหน้าที่ต่างกันของระดับที่ขั้น j

$\beta_{j2} = 0$ หมายถึง ไม่เกิดการทำหน้าที่ต่างกันของขั้นที่ขั้น j

$\beta_{j2} > 0$ หมายถึง เกิดการทำหน้าที่ต่างกันของขั้นเข้าข้างกลุ่มอ้างอิง

$\beta_{j2} < 0$ หมายถึง เกิดการทำหน้าที่ต่างกันของขั้นเข้าข้างกลุ่มเปรียบเทียบ

2.6 การใช้ผลการทำหน้าที่ต่างกันเป็นขั้นเพื่อทดสอบการทำหน้าที่ต่างกันของข้อสอบ

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสามารถทดสอบจากผลการทำหน้าที่ต่างกันเป็นขั้น ถ้าสมมติฐานว่างของการไม่เกิดการทำหน้าที่ต่างกันเป็นขั้นยอมรับในทุกๆ ขั้น ดังนั้นสมมติฐานว่างของการไม่เกิดการทำหน้าที่ต่างกันของข้อสอบอาจจะถูกยอมรับได้เช่นกัน แต่ถ้าสมมติฐานว่างของการไม่เกิดการทำหน้าที่ต่างกันเป็นขั้นถูกปฏิเสธในขั้นใดขั้นหนึ่ง ดังนั้นสมมติฐานว่างของการไม่เกิดการทำหน้าที่ต่างกันของข้อสอบสามารถถูกปฏิเสธตามไปด้วย ซึ่งวิธีการดังกล่าวนี้เรียกว่า การทดสอบระดับขั้นพร้อมๆ กันของการทดสอบการทำหน้าที่ต่างกันของข้อสอบ

(simultaneous step-level test of DIF) (Penfield, 2007a; Penfield, Gattamorta, & Childs, 2009)

การทดสอบระดับขั้นพร้อมๆ กันของการทดสอบการทำหน้าที่ต่างกันของข้อสอบ (simultaneous step-level (SSL) แสดงให้เห็นอำนาจการทดสอบมากกว่าวิธีอื่นๆ ที่ใช้ในการทดสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อขนาด และ/หรือ สัญลักษณ์ของขนาดการทดสอบการทำหน้าที่ต่างกันเป็นขั้นมีความหลากหลายข้ามระดับ (Penfield, Gattamorta, & Childs, 2009) ทั้งนี้การทดสอบระดับขั้นพร้อมๆ กันของการทดสอบการทำหน้าที่ต่างกันของข้อสอบ (SSL) มีความสัมพันธ์กับการทดสอบ net DIF และ global DIF โดยที่ การทดสอบ net DIF อยู่บนพื้นฐานการรวมกันของเครื่องหมายผลการทำหน้าที่ต่างกันเป็นขั้นทุกๆ ขั้น และไม่ไวต่อผลการทำหน้าที่ต่างกันเป็นขั้นแบบลู่ออก (divergent) และเป็นการทดสอบที่ให้ความสำคัญกับความแตกต่างระหว่างกลุ่ม (between-group differences) ในค่าคาดหวัง (expected value) ของตัวแปรตอบสนองที่มีการให้คะแนนพหุวิภาคส่วน การทดสอบ global DIF พิจารณาจากผลการทำหน้าที่ต่างกันเป็นขั้นไม่คิดเครื่องหมายทุกๆ ขั้น ดังนั้นจึงไวต่อผลการทำหน้าที่ต่างกันเป็นขั้นแบบลู่ออก โดยแนวคิดของ global DIF เป็นการทดสอบที่ให้ความสำคัญกับความเป็นอิสระของสมาชิกกลุ่ม (dependence of group membership) และตัวแปรตอบสนองที่มีการให้คะแนนหลายค่า (Penfield, 2010; Penfield, Alvarez, & Lee, 2008; Penfield, Gattamorta, & Childs, 2009)

ดังนั้นการทดสอบระดับขั้นพร้อมๆ กันของการทดสอบการทำหน้าที่ต่างกันของข้อสอบ (SSL) จึงเป็นการทดสอบ global DIF เนื่องจากไม่ได้มีการแยกผลการทำหน้าที่ต่างกันเป็นขั้นคิดเครื่องหมายในทุกๆ ขั้น ซึ่งในการทดสอบมีความไวมากกว่าการทดสอบ net DIF เมื่อผลการทำหน้าที่ต่างกันเป็นขั้นเป็นแบบลู่ออก (divergent) หรือมีขนาดใหญ่ รายละเอียดนิยามและลักษณะการเกิด net DIF และ global DIF สรุปรายละเอียดตามตารางที่ 8

ตารางที่ 8 สรุป net DIF และ global DIF

ลักษณะการเกิด	net DIF	global DIF
นิยาม	เป็นการทดสอบที่ให้ความสำคัญกับความแตกต่างระหว่างกลุ่ม (between-group differences) ในค่าคาดหวัง (expected value) ของตัวแปรตอบสนองที่มีการให้คะแนนหลายค่า	เป็นการทดสอบที่ให้ความสำคัญกับความเป็นอิสระของสมาชิกกลุ่ม (dependence of group membership) และตัวแปรตอบสนองที่มีการให้คะแนนหลายค่า
เครื่องหมาย	การรวมกันของเครื่องหมายผลการทำหน้าที่ต่างกันเป็นขั้นทุกๆ ขั้น และไม่ไวต่อผล การทำหน้าที่ต่างกันเป็นขั้นแบบลู่ออก (divergent)	การทำหน้าที่ต่างกันเป็นขั้นไม่คิดเครื่องหมายทุกๆ ขั้น ดังนั้นจึงไวต่อผลการทำหน้าที่ต่างกันเป็นขั้นแบบ ลู่ออก (divergent)
ลักษณะการทำหน้าที่ต่างกันเป็นขั้น	$\omega_j = b_{jF} - b_{jR}$	$\Omega = [\omega_1, \ldots, \omega_2]$

หมายเหตุ ω_j หมายถึง ผลการทำหน้าที่ต่างกันเป็นขั้น (the DSF effect)

Ω หมายถึง เวกเตอร์ของผลการทำหน้าที่ต่างกันเป็นขั้น

2.7 งานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันเป็นขั้น

จากการศึกษางานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันเป็นขั้น พบว่า ในปัจจุบันการศึกษาการทำหน้าที่ต่างกันเป็นขั้นเป็นแนวคิดที่นักวิชาการกำลังให้ความสนใจศึกษาเพิ่มขึ้นและหลากหลายมิติ ทั้งนี้ผู้วิจัยได้จำแนกออกเป็น 4 กลุ่ม ได้แก่ กลุ่มที่ 1 ศึกษาตัวประมาณค่าการทำหน้าที่ต่างกันเป็นขั้น (Penfield, 2007b; Penfield, 2008; Gattamorta & Penfield, 2012) กลุ่มที่ 2 ศึกษาความไม่แปรเปลี่ยนของโมเดล (Penfield, Mayer, & Wolfe, 2008) กลุ่มที่ 3 ศึกษาการทำหน้าที่ต่างกันเป็นขั้นกับข้อมูลจริง (Penfield, Gattamorta, & Childs, 2009; Miller, Chahine, & Childs, 2010) และกลุ่มที่ 4 ศึกษาการทำหน้าที่ต่างกันเป็นขั้นตามลักษณะการทำหน้าที่ต่างกันของข้อสอบ (Penfield, 2010b)

กลุ่มที่ศึกษาตัวประมาณค่าการทำหน้าที่ต่างกันเป็นขั้น ได้แก่ Penfield (2007b) ซึ่งศึกษาการทำหน้าที่ต่างกันของข้อสอบโดยใช้ตัวประมาณค่าอัตราส่วนเลขออด (common odds ratio) โดยทั่วไปแล้วการใช้ดัชนีระดับข้อสอบเดี่ยว (single item-level index) สารสนเทศของข้อสอบที่มีการตรวจให้คะแนนพหุวิภาคอาจมีความคลาดเคลื่อนเมื่อขนาด (magnitude) และทิศทาง (direction) มีการเปลี่ยนแปลงข้ามระดับในข้อสอบ การวิจัยครั้งนี้เป็นการทดสอบความไม่แปรเปลี่ยนในการวัดรูปแบบข้อสอบที่มีการตรวจให้คะแนนพหุวิภาคเป็นการทดสอบความไม่แปรเปลี่ยนในแต่ละขั้น โดยใช้กรอบแนวคิดตามวิธีการต่างกันเป็นขั้นและทดสอบผ่านตัวประมาณ

ค่าการทำหน้าที่ต่างกันเป็นขั้นแบบนันพาราเมทริก (nonparametric DSF estimator) ผ่านการศึกษาโดยใช้การจำลองข้อมูล จากการวิจัยพบว่า ระดับการทำหน้าที่ต่างกันเป็นขั้น (level of DSF) มี การแปรเปลี่ยนขนาดหรือเครื่องหมายข้ามระดับที่อยู่ภายในตัวเลือกของข้อสอบ วิธีการทำหน้าที่ต่างกันเป็นขั้นในการทดสอบไม่แปรเปลี่ยนในการวัดจะให้อำนาจในการทดสอบและ ความแม่นยำในการทดสอบสูงมากกว่าตัวประมาณค่าการทำหน้าที่ต่างกันระดับข้อสอบ (item-level DIF estimators)

ในปีต่อมา Penfield (2008) ได้ทำการศึกษาสามระดับของตัวประมาณค่าผลการทำหน้าที่ต่างเป็นขั้น (DSF) ในตัวแปรประเภทนันพาราเมทริก เป็นการทดสอบความไม่แปรเปลี่ยนการวัดในข้อสอบที่มีการตรวจให้คะแนนพหุวิภาคโดยที่ขนาด (magnitude) และ (sign) อาจจะเปลี่ยนแปลงระดับตัวเลือกในข้อสอบที่มีการตรวจให้คะแนนพหุวิภาค แนวคิดดังกล่าวถูกอ้างอิงจากวิธีการทำหน้าที่ต่างกันเป็นขั้น (DSF) ทั้งนี้ตัวประมาณค่าที่ใช้ในการศึกษาเป็นตัวประมาณค่าที่ศึกษาจาก 3 โมเดล คือ the graded response model (GRM) the continuation ratio model (CRM) และ the generalized partial credit model (GPCM) โดยใช้วิธีการจำลองข้อมูลและเปรียบเทียบค่าสถิติจากตัวประมาณค่าผลการทำการทำหน้าที่ต่างกันเป็นขั้น (DSF) 3 ขั้น จากผลการวิจัยชี้ให้เห็นว่าโมเดล GRM มีระดับความคงที่สูงสุดและมีความลำเอียงในทุกๆ ขั้นน้อยที่สุด

ต่อมา Gattamorta และ Penfield (2012) ได้ศึกษาเปรียบเทียบการกำหนดหมวดหมู่แบบใกล้เคียงกัน (adjacent categories) และตัวประมาณค่าผลการทำหน้าที่ต่างกันเป็นขั้นแบบสะสมในการศึกษาเป็นการวัดความไม่แปรเปลี่ยนในข้อสอบที่มีการตรวจให้คะแนนพหุวิภาคพบว่า แนวคิดเดิมในการตรวจสอบการทำหน้าที่ต่างกันเป็นขั้นมี 2 แนวคิด คือ แนวคิดการกำหนดหมวดหมู่แบบใกล้เคียงกัน (adjacent categories) และแนวคิดแบบสะสม (cumulative approach) ซึ่งยังมีข้อจำกัดเกี่ยวกับงานวิจัยทั้งสองแนวคิดในการเปรียบเทียบการทำหน้าที่ต่างกันเป็นขั้น โดยเฉพาะการประยุกต์ใช้กับข้อมูลจริง ดังนั้นในการศึกษาครั้งนี้จึงเป็นการประเมินผลการวิเคราะห์การทำหน้าที่ต่างกันเป็นขั้นของทั้งสองแนวคิด เพื่อที่จะระบุกรณีที่ไม่เกิดผลการทำหน้าที่ต่างกันเป็นขั้นและเมื่อเกิดผลการทำหน้าที่ต่างกันเป็นขั้นแล้วทั้งสองแนวคิดมีข้อสรุปที่แตกต่างกันอย่างไร

กลุ่มที่ศึกษาความไม่แปรเปลี่ยนของโมเดลได้แก่ Penfield, Mayer, และ Wolfe (2008) ได้ศึกษาวิธีการประเมินความไม่แปรเปลี่ยนของข้อสอบ ระดับขั้น และ Threshold ในข้อสอบที่มีการตรวจให้คะแนนพหุวิภาคตาม partial credit model (PCM) โดยใช้วิธีเชิงพาราเมทริกตารางเลขจรสำหรับการประเมินความไม่แปรเปลี่ยนทั้ง 3 ประเภท จากการวิจัย พบว่า ผลจากตัวประมาณความไม่แปรเปลี่ยน (invariance effect estimates observed) จากทั้งวิธีเชิงพาราเมทริกและตารางเลขจรมีความคงที่และสอดคล้องเชิงทฤษฎี

กลุ่มที่ศึกษาการทำหน้าที่ต่างกันเป็นขั้นกับข้อมูลจริง ได้แก่ Penfield, Gattamorta, และ Childs (2009) ได้ใช้การทำหน้าที่ต่างกันเป็นขั้น (DSF) ใช้ในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบในข้อสอบที่มีการตรวจให้คะแนนพหุวิภาคซึ่งวิธีการแบบเดิมนั้นเป็นเพียงการตรวจที่ให้ผลดัชนีระดับข้อสอบของการทำหน้าที่ต่างกันของข้อสอบ (single item-level index of DIF) และไม่ได้ให้ข้อมูลที่เกี่ยวข้องกับคะแนนแต่ละระดับที่สามารถบอกให้เห็นการทำหน้าที่ต่างกันของข้อสอบได้ ด้วยเหตุนี้ยังเป็นจำกัดในวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในปัจจุบัน อย่างไรก็ตาม วิธีการทำหน้าที่ต่างกันเป็นขั้นถูกนำเสนอขึ้น โดยที่ความไม่แปรเปลี่ยนในการวัด (measurement invariance) ถูกทดสอบภายในแต่ละระดับที่อยู่ภายใต้ตัวแปรตอบสนองเชิงกลุ่ม (polytomous response variable) จากวิธีการทดสอบดังกล่าวแสดงให้คุณค่าของข้อมูลสารสนเทศที่เกี่ยวข้องกับธรรมชาติของการทำหน้าที่ต่างกันของข้อสอบ (DIF) เช่น การเกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF) เป็นผลในระดับข้อสอบ (item-level effects) หรือผลในระดับคะแนน (specific score levels) ตำแหน่งของการเกิดการทำหน้าที่ต่างกันของข้อสอบ (The location of the DIF effect) เช่น ระดับคะแนนที่ชัดเจนที่แสดงผลการทำหน้าที่ต่างกันของข้อสอบ และศักยภาพของสาเหตุการทำหน้าที่ต่างกันของข้อสอบ (the potential causes of a DIF effect) เช่นลักษณะข้อสอบหรือความลำเอียงในการคุมสอบ และจากการวิจัยได้เสนอตัวอย่างข้อมูลที่ได้จากการทดสอบจริงเพื่อให้เข้าใจวิธีการการทำหน้าที่ต่างกันเป็นขั้น (DSF) มากยิ่งขึ้น

นอกจากกลุ่มของ Penfield, Gattamorta, และ Childs (2009) ต่อมา Miller, Chahine, และ Childs (2010) ยังได้ศึกษาการทำหน้าที่ต่างกันของข้อสอบ (DIF) และการทำหน้าที่ต่างกันเป็นขั้น (DSF) เพื่อใช้ตรวจสอบความยากของข้อสอบ (item difficulty) ที่สัมพันธ์กับประสบการณ์ของผู้เข้าสอบ เช่น การสอนของครูในนักเรียน ที่สัมพันธ์กับความรู้ ทักษะ และความสามารถการสอบ การวิเคราะห์ดังกล่าวเป็นการเปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบโดยทั่วไปกับการทำหน้าที่ต่างกันเป็นขั้น (DSF) โดยที่คุณลักษณะด้านเพศ ภาษา หรือความรู้เชิงวัฒนธรรม ไม่ควรที่จะความสัมพันธ์กัน การทดสอบใช้ข้อมูลมาวิเคราะห์ทั้ง 2 รูปแบบ ในบริบทการประเมินความรู้นักเรียนระดับเกรด 9 วิชาคณิตศาสตร์ของเมือง Ontario และเปรียบเทียบกลุ่มนักเรียนที่ถูกนิยามจากการการสอนของครู โดยใช้การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ตามวิธีแมนเทล- แฮนส์เซล (MH) วิธี standardized liu- agresti cumulative common log-odds ratio และวิธี standardized Cox’s noncentrality parameter จากผลการวิจัย พบว่า การปรากฏของข้อสอบที่มีการทำหน้าที่ต่างกันของข้อสอบในระดับกลางถึงระดับใหญ่ถูกทดสอบตามมาภายหลัง โดยวิธีการทำหน้าที่ต่างกันของระดับจะทำให้บอกถึงการพัฒนาข้อสอบและศักยภาพในการสอนของครูได้

นอกจากนี้ กลุ่มที่ศึกษาการทำหน้าที่ต่างกันเป็นขั้นตามลักษณะการทำหน้าที่ต่างกันของข้อสอบ ได้แก่ Penfield (2010b) ซึ่งอธิบายการการตรวจสอบการทำหน้าที่ต่างกันแบบตัดกันของข้อสอบ (CDIF) ในข้อสอบที่มีการตรวจให้คะแนนพหุวิภาค โดยใช้ผลการจากทำหน้าที่ต่างกันเป็นขั้น (differential step functioning effects) เนื่องจากปัจจุบันการตรวจสอบการทำหน้าที่ต่างกันแบบตัดกันในข้อสอบยังมีปัญหาในเรื่องสถิติทดสอบหลายตัวที่ถูกพัฒนาสำหรับการประเมินการทำหน้าที่ต่างกันของข้อสอบ (DIF) เนื่องจากความแตกต่างระหว่างกลุ่มทั้งที่เป็นไปเงื่อนไขแบบบวก (positive) และเงื่อนไขแบบลบ (negative) ในคะแนนที่คาดหวังของข้อสอบ (expected item score) สามารถปฏิเสธผลที่ขาดความแตกต่างระหว่างกลุ่มเมื่อมีการรวมข้ามขนาดของคุณลักษณะแฝง (latent trait continuum) การวิจัยครั้งนี้เป็นการศึกษาจากการจำลองข้อมูล การตรวจสอบการทำหน้าที่ต่างกันแบบตัดกันในข้อสอบสามารถอธิบายโดยใช้ผลจากการทำหน้าที่ต่างกันเป็นขั้น (differential step functioning effects) ที่มีการแปรเปลี่ยนสัญลักษณ์ (sign) ข้ามระดับ

จากการศึกษางานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันเป็นขั้น สามารถสรุปผล การวิเคราะห์องค์ความรู้ที่ได้จากงานวิจัยที่ศึกษารายละเอียดตามตารางที่ 9 ดังนี้

ตารางที่ 9 สรุปผลองค์ความรู้จากงานวิจัยการทำหน้าที่ต่างกันเป็นขั้น

ผู้วิจัย	วิธีการดำเนินงาน	องค์ความรู้ในการศึกษา
Penfield (2007b)	ศึกษาการทำหน้าที่ต่างกันในข้อสอบโดยใช้ตัวประมาณค่าอัตราส่วนเลขออด (common odds ratio)	ใช้ตัวประมาณค่าอัตราส่วนเลขออด (common odds ratio) ในการตรวจสอบการทำหน้าที่ต่างกันในข้อสอบที่มีการตรวจให้คะแนนหลายค่า
Penfield (2008)	ศึกษาสามระดับของตัวประมาณค่าผลการทำหน้าที่ต่างเป็นขั้น (DSF) ในตัวแปรประเภทนันพาราเมทริกซ์ ใน 3 โมเดล คือโมเดล GRM โมเดล CRM และโมเดล GPCM	ตัวประมาณค่าผลการทำการทำหน้าที่ต่างกันเป็นขั้น (DSF) 3 ระดับ ชี้ให้เห็นว่าโมเดล GRM มีระดับความคงที่สูงสุดและมีความลำเอียงทุกๆ ระดับน้อยที่สุด

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 9 สรุปผลองค์ความรู้จากงานวิจัยการทำหน้าที่ต่างกันเป็นขั้น (ต่อ)

ผู้วิจัย	วิธีการดำเนินงาน	องค์ความรู้ในการศึกษา
Penfield, Mayer, และ Wolfe (2008)	ศึกษาวิธีการประเมินความไม่แปรเปลี่ยนของข้อสอบ ระดับขั้น และ Threshold ในข้อสอบที่มีการตรวจให้คะแนนพหุวิภาคตามโมเดล PCM	ตัวประมาณค่าความไม่แปรเปลี่ยน (invariance effect estimates observed) จากทั้งวิธีเชิงพาราเมทริก และตารางเลขจรมีความคงที่สอดคล้องเชิงทฤษฎี
Penfield, Gattamorta, และ Childs (2009)	ศึกษาการทำหน้าที่ต่างกันเป็นขั้นใช้ในการวิเคราะห์สำหรับการทำหน้าที่ต่างกันของข้อมูลจริงในข้อสอบที่มีการตรวจให้คะแนนพหุวิภาค	องค์ความรู้โมเดลการทำหน้าต่างกันเป็นขั้นกับข้อสอบจริงที่มีการตรวจให้คะแนนพหุวิภาค
Penfield (2010b)	ศึกษาการการตรวจสอบการทำหน้าที่ต่างกันแบบตัดกันในข้อสอบที่มีการตรวจให้คะแนนพหุวิภาค โดยใช้ผลการจากทำหน้าที่ต่างกันเป็นขั้น (differential step functioning effects)	การตรวจสอบการทำหน้าที่ต่างกันแบบตัดกันในข้อสอบที่มีการตรวจให้คะแนนพหุวิภาค สามารถใช้วิธีจากการทำหน้าที่ต่างกันเป็นขั้นในการอธิบาย
Miller, Chahine, และ Childs (2010)	ศึกษาการทำหน้าที่ต่างกันของข้อสอบและการทำหน้าที่ต่างกันเป็นขั้นเพื่อตรวจสอบความยากของข้อสอบ (item difficulty) ที่สัมพันธ์กับประสบการณ์ของผู้เข้าสอบ	วิธีการทำหน้าที่ต่างกันของระดับเมื่อข้อสอบที่มีผลการทำหน้าที่ต่างกันเป็นขั้นระดับกลางถึง ระดับใหญ่ สามารถนำไปศึกษาต่อในการพัฒนาข้อสอบ และศักยภาพในการสอนของครู ได้
Gattamorta และ Penfield (2012)	ศึกษาเปรียบเทียบการกำหนดหมวดหมู่แบบใกล้เคียงกัน (adjacent categories) และตัวประมาณค่าผลการทำหน้าที่ต่างกันเป็นขั้นแบบสะสม	วิเคราะห์การทำหน้าที่ต่างกันเป็นขั้นของทั้งแนวคิดการกำหนดหมวดหมู่แบบใกล้เคียงกัน (adjacent categories) และตัวประมาณค่าผลการทำหน้าที่ต่างกันเป็นขั้นแบบสะสม

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

การทำหน้าที่ต่างกันเป็นขั้น (differential step functioning: DSF) เป็นวิธีที่ใช้ในการตรวจสอบความแตกต่างระหว่างกลุ่ม (between-group differences) ในแต่ละขั้น (step) ของข้อสอบที่มีการตรวจให้คะแนนพหุวิภาค การอธิบายขั้นตามการทำหน้าที่ต่างกันเป็นขั้นสามารถอธิบายได้ตามทฤษฎีตอบสนองข้อสอบโลจิสติก 2 พารามิเตอร์ ซึ่งพารามิเตอร์ค่าอำนาจจำแนกมีลักษณะคงที่ ส่วนพารามิเตอร์ความยากขึ้นอยู่กับลักษณะกลุ่ม นอกจากนี้รูปแบบการทำหน้าที่ต่างกันเป็นขั้นที่เชื่อมโยงผลการทำหน้าที่ต่างกันของข้อสอบสามารถแบ่งได้เป็น 2 มิติ มิติที่หนึ่งอธิบายตามการจัดกลุ่มภายในรูปแบบการทำหน้าที่ต่างกันเป็นขั้น ซึ่งมีลักษณะกระจาย (pervasive) และ ลักษณะไม่กระจาย (nonpervasive) และมิติที่สองผลการทำหน้าที่ต่างกันเป็นขั้นมีลักษณะคงที่ (constant) ลักษณะลู่เข้า (convergent) และลักษณะลู่ออก (divergent)

วิธีการแบ่งหมวดหมู่ผลการทำหน้าที่ต่างกันเป็นขั้นใช้ อัตราส่วนลอกเลขออดร่วมขั้นเป็นระดับ (the step-level common log-odds ratio) สำหรับขั้นที่ j ใดๆ ($\hat{\lambda}_j$) แบ่งได้เป็น ขนาดเล็ก ขนาดกลาง และขนาดใหญ่ นอกจากนี้วิธีการเชิงสถิติที่ใช้ในการตรวจสอบการทำหน้าที่ต่างกันเป็นขั้น ประกอบด้วย วิธีทฤษฎีการตอบสนองข้อสอบ (IRT approach) วิธีอัตราส่วนเลขออด (odds ratio approach) และวิธีถดถอยโลจิสติก (logistic regression approach) ซึ่งเป็นแนวคิดแบบ cumulative approach

การใช้กรอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) และการทำหน้าที่ต่างกันเป็นขั้น (DSF) จำเป็นที่จะต้องคำนึงถึงจุดแข็งและจุดอ่อนของแต่ละกรอบแนวคิด ทั้งนี้กรอบแนวคิดการทำหน้าที่ต่างกันของข้อสอบ (DIF) มีจุดอ่อนที่ไม่ได้ให้สารสนเทศที่เป็นคะแนนในแต่ละขั้นที่เกี่ยวข้องกับอิทธิพลการทำหน้าที่ต่างกันของข้อสอบ ในขณะที่กรอบแนวคิดการทำหน้าที่ต่างกันเป็นขั้น (DSF) ให้สารสนเทศที่ครอบคลุมคะแนนในแต่ละขั้นซึ่งสะท้อนผลการทำหน้าที่ต่างกันของข้อสอบในระดับข้อสอบได้ ถึงแม้ว่ากรอบแนวคิดการทำหน้าที่ต่างกันของข้อสอบจะมีข้อจำกัดทางด้านสารสนเทศของข้อสอบ แต่แนวคิดการทำหน้าที่ต่างกันของข้อสอบยังมีอำนาจในการตรวจสอบมากกว่าการทดสอบการทำหน้าที่ต่างกันเป็นขั้นเพราะว่าการทำหน้าที่ต่างกันของข้อสอบเป็นการวิเคราะห์องค์รวมของทุกๆ ขั้น พร้อมกัน อาจกล่าวได้ว่ากรอบการทำหน้าที่ต่างกันของข้อสอบมีความไวต่อการตรวจสอบการวัดที่ไม่มีความสอดคล้องกัน และกรอบการทำหน้าที่ต่างกันเป็นขั้นให้สารสนเทศในการตรวจสอบว่าการวัดความไม่สอดคล้องกัน (Penfield, Gattamorta, & Childs, 2009)

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง พบว่า การศึกษาการทำหน้าที่ต่างกันเป็นขั้นเป็นแนวคิดการทำหน้าที่ต่างกันของข้อสอบของนักวิชาการทางด้านวัดผลที่อธิบายการเกิดการทำหน้าที่ต่างกันของข้อสอบเป็นระดับขั้นข้อสอบที่มีการตรวจให้คะแนนพหุวิภาค (polytomous items) ซึ่งมีทั้งการศึกษาตัวประมาณค่าการทำหน้าที่ต่างกันเป็นขั้นและศึกษาตามลักษณะการทำหน้าที่ต่างกันแบบตัดกันของข้อสอบ อย่างไรก็ตาม จากการศึกษางานวิจัยที่เกี่ยวข้องการทำหน้าที่

ต่างกันเป็นขั้น พบว่า ยังมีงานวิจัยจำนวนน้อยที่ให้ความสนใจประเด็นการทำหน้าที่ต่างกันเป็นขั้น มีเพียงงานวิจัยของ Penfield, Gattamorta, และ Childs (2009) และ Miller, Chahine, และ Childs (2010) ที่นำองค์ความรู้ทางด้านการทำหน้าที่ต่างกันเป็นขั้นมาศึกษากับข้อมูลจริง เพื่อใช้อธิบายความสัมพันธ์การเกิดการทำหน้าที่ต่างกันของข้อสอบและการเกิดการทำหน้าที่ต่างกันเป็นขั้น พร้อมกับการพิจารณาขนาดอิทธิพลในการเกิดการทำหน้าที่ต่างกันเป็นขั้นในข้อสอบที่มีการตรวจให้คะแนนพหุวิภาค ดังนั้นเพื่อเป็นการพัฒนาองค์ความรู้ในการตรวจสอบการทำหน้าที่ต่างเป็นขั้นในบริบททางการศึกษา ผู้วิจัยจึงมีความสนใจที่จะนำประเด็นการทำหน้าที่ต่างกันเป็นขั้นมาศึกษาเพิ่มเติมและพัฒนาองค์ความรู้ดังกล่าวให้มากขึ้น อันจะนำมาซึ่งประโยชน์ทางด้านการพัฒนาทางการวัดผลและทางด้านการศึกษาของประเทศไทยต่อไป

ตอนที่ 3 โครงการประเมินผลนักเรียนนานาชาติ (programme for international student assessment: PISA)

โครงการประเมินผลนักเรียนนานาชาติหรือเรียกว่า โครงการ PISA (programme for international student assessment: PISA) เป็นโครงการศึกษาสำรวจความรู้และทักษะของนักเรียนกลุ่มอายุ 15 ปี ในประเทศที่เป็นสมาชิกองค์กรเพื่อความร่วมมือและพัฒนาทางเศรษฐกิจ (OECD) และประเทศร่วมโครงการที่ไม่ใช้สมาชิก ซึ่งหนึ่งในประเทศที่ไม่ใช่สมาชิกคือประเทศไทย โดยวัตถุประสงค์หลักของโครงการดังกล่าวเพื่อให้ข้อมูลกับประเทศสมาชิกว่าได้เตรียมความพร้อมของประชาชนให้มีศักยภาพสำหรับการแข่งขันในประชาชมโลกเพียงพอหรือไม่อย่างไร และเพื่อหาตัวชี้วัดคุณภาพการศึกษาในระดับนโยบาย โดยตัวชี้วัดที่สำคัญในการประเมินได้แก่ การรู้เรื่องคณิตศาสตร์ (mathematics literacy) การรู้เรื่องการอ่าน (reading literacy) และ การรู้เรื่องวิทยาศาสตร์ (scientific literacy)

เนื่องจาก PISA มีวัตถุประสงค์เพื่อชี้บอกแนวโน้มการเปลี่ยนแปลงของผลการศึกษาของประเทศในโครงการ PISA จึงมีการประเมินผลทุกๆ สามปี แต่การประเมินผลแต่ละครั้งจะให้น้ำหนักสำหรับกับการประเมินต่างกัน โดยการประเมินผลถูกกำหนดเป็นรอบ (cycle) รอบแรก เริ่มต้นมาตั้งแต่ PISA 2000 จึงครบรอบหนึ่งใน PISA 2006 และเริ่มต้นรอบที่สอง (second cycle) ในปี 2009 ซึ่งเป็นโครงการที่ให้น้ำหนักกับการอ่านเช่นเดียวกับ PISA 2000 ในรอบแรก ทั้งนี้ การประเมินความรู้และทักษะที่ PISA นิยามว่าเป็นการรู้เรื่อง (literacy) มีจุดเน้นให้น้ำหนัก การประเมินในแต่ละครั้งต่างกัน PISA 2000 ให้น้ำหนักสำคัญกับ การรู้เรื่องการอ่าน (reading literacy) PISA 2003 ให้น้ำหนักกับการรู้เรื่องคณิตศาสตร์ (mathematical literacy) และ PISA 2006 ให้น้ำหนักกับการู้เรื่องวิทยาศาสตร์ (scientific literacy) ส่วนรอบที่สอง PISA 2009 มี

น้ำหนักข้อสอบเช่นเดียวกับ PISA 2000 ส่วน PISA 2012 มีน้ำหนักข้อสอบเช่นเดียวกับ PISA 2003 และ PISA 2015 มีน้ำหนักข้อสอบเช่นเดียวกับ PISA 2006 (สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี, 2554) รายละเอียดตามตารางที่ 10

ตารางที่ 10 รอบการประเมินโครงการประเมินผลนักเรียนนานาชาติ

รอบประเมิน	รอบที่หนึ่ง (first cycle)			รอบที่สอง (second cycle)		
	2000	2003	2006	2009	2012	2015
วิชาหลัก	การอ่าน	คณิตศาสตร์	วิทยาศาสตร์	การอ่าน	คณิตศาสตร์	วิทยาศาสตร์
วิชารอง	คณิตศาสตร์	วิทยาศาสตร์	การอ่าน	คณิตศาสตร์	วิทยาศาสตร์	การอ่าน
	วิทยาศาสตร์	การอ่าน	คณิตศาสตร์	วิทยาศาสตร์	การอ่าน	คณิตศาสตร์

3.1 เครื่องมือที่ใช้ในโครงการประเมินผลนักเรียนนานาชาติ

เครื่องมือที่ใช้ในโครงการประเมินผลนักเรียนนานาชาติ หรือ PISA ประกอบด้วยแบบสอบวัดการรู้เรื่องทั้งสามด้าน คือ การรู้เรื่องวิทยาศาสตร์ (scientific literacy) การรู้เรื่องการอ่าน (reading literacy) และการรู้เรื่องคณิตศาสตร์ (mathematics literacy) โดยมีรายละเอียดแต่ละด้านดังนี้

3.1.1 การรู้เรื่องวิทยาศาสตร์ (scientific literacy)

โครงการ PISA ได้ให้นิยามการรู้เรื่องวิทยาศาสตร์ คือ บุคคลที่มีความรู้ทางวิทยาศาสตร์และใช้ความรู้นั้นในการระบุประเด็นปัญหา (คำถาม) เพื่อหาความรู้ใหม่อธิบายปรากฏการณ์ และตัดสินใจเรื่องที่เกี่ยวกับวิทยาศาสตร์บนพื้นฐานของประจักษ์พยานวิทยาศาสตร์ มีความเข้าใจลักษณะพิเศษของวิทยาศาสตร์ซึ่งเป็นรูปแบบหนึ่งของการหาความรู้ของมนุษย์ มีความตระหนักว่าวิทยาศาสตร์และเทคโนโลยี ที่ทำให้เกิดวัสดุใหม่ ตลอดจนสิ่งแวดล้อมทางปัญญาและวัฒนธรรมใหม่ของมนุษยชาติได้อย่างไร และมีความตั้งใจและเต็มใจที่จะมีส่วนร่วมงานวิทยาศาสตร์และมีความคิดทางวิทยาศาสตร์ อย่างพลเมืองที่มีความคิดรับผิดชอบ นอกจากนี้การประเมินการรู้เรื่องวิทยาศาสตร์ใน PISA ประกอบด้วยส่วนที่มีความสัมพันธ์ซึ่งกันและกันสี่ประการ คือ 1) สถานการณ์และบริบท การับรู้ถึงสถานการณ์ในชีวิตที่เกี่ยวข้องกับวิทยาศาสตร์และเทคโนโลยี 2) ความรู้ ความเข้าใจในโลกธรรมชาติ บนพื้นฐานของวิทยาศาสตร์และเทคโนโลยี รวมทั้งความรู้วิทยาศาสตร์ และความรู้เกี่ยวกับวิทยาศาสตร์ 3) สมรรถนะทางวิทยาศาสตร์ การใช้ความรู้วิทยาศาสตร์ในการระบุประเด็นทางวิทยาศาสตร์ (identifying scientific issues) การอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ (explain phenomena scientifically) และการใช้ประจักษ์

พยานทางวิทยาศาสตร์ (using scientific evidence) และ 4) เจตคติเชิงวิทยาศาสตร์ ได้แก่ การแสดงการตอบสนองต่อวิทยาศาสตร์ด้วยความสนใจ สนับสนุน การสืบหาความรู้วิทยาศาสตร์ และแสดงความรับผิดชอบต่อสิ่งต่างๆ เช่น ในประเด็นของทรัพยากรธรรมชาติและสิ่งแวดล้อม เป็นต้น

3.1.2 การรู้เรื่องการอ่าน (reading literacy)

โครงการ PISA ได้ให้นิยามการรู้เรื่องการอ่าน หมายถึง การรู้เรื่องการอ่านเป็นความเข้าใจ (understanding) การใช้ (using) การสะท้อน (reflecting) และความรักและผูกพันกับการอ่าน (engaging) ในถ้อยความที่เป็นข้อเขียน (written texts) ที่ได้อ่าน เพื่อไปบรรลุเป้าหมายของแต่ละคน เพื่อพัฒนาความรู้และศักยภาพของตนและเพื่อมีส่วนร่วมในกระบวนการของสังคม นอกจากนี้การประเมินการรู้เรื่องการอ่านในโครงการ PISA ตั้งอยู่บนพื้นฐาน 3 ประการด้วยกัน ได้แก่ 1) สถานการณ์ เป็นบริบทหรือจุดประสงค์ของการอ่านที่เกิดขึ้น 2) ถ้อยความหรือเนื้อความเป็นสิ่งที่เขียนหรือเรียบเรียงให้อ่านให้ครอบคลุมพิสัยความยากง่ายของสื่อหรือวัสดุที่อ่าน และ 3) กลยุทธ์การอ่านเป็นกลยุทธ์เชิงความคิดที่เป็นตัวกำหนดว่าผู้อ่านอ่านในลักษณะใด

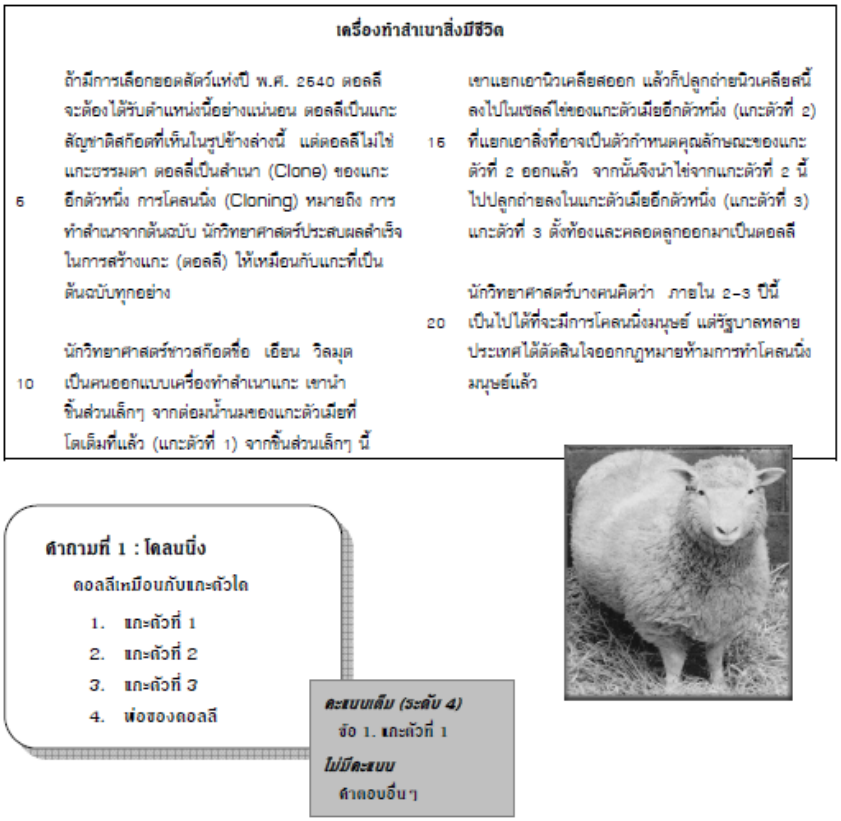
3.1.3 การรู้เรื่องคณิตศาสตร์ (mathematics literacy)

โครงการประเมินผล PISA ได้ให้นิยามการรู้เรื่องคณิตศาสตร์ หมายถึง สมรรถนะของบุคคลในการที่จะบ่งบอกและเข้าใจบทบาทของคณิตศาสตร์ที่มีในโลก เพื่อให้สามารถตัดสินใจบนพื้นฐานความรู้ที่เข้มแข็ง และเพื่อใช้และผูกพันกับคณิตศาสตร์ ที่จะตอบสนองความจำเป็นต่อชีวิตของบุคคล ในอันที่จะเป็นพลเมืองที่มีความคิด มีความห่วงใย และสร้างสรรค์สังคม นอกจากนี้การประเมินการรู้เรื่องคณิตศาสตร์ใน PISA ครอบคลุมองค์ประกอบ 3 ด้าน ได้แก่ 1) สถานการณ์หรือบริบทที่ปัญหานั้นตั้งอยู่ (situation on context) 2) เนื้อหาคณิตศาสตร์ (mathematical content) ที่ต้องนำมาใช้ในการแก้ปัญหา และ 3) สมรรถนะทางคณิตศาสตร์ของนักเรียนที่ควรได้รับการกระตุ้นเร้าให้สามารถเชื่อมต่อกับโลกจริงๆ ที่ปัญหานั้นๆ เกิดขึ้นโดยใช้คณิตศาสตร์ และให้สามารถแก้ปัญหาได้โดยใช้คณิตศาสตร์นั้นๆ

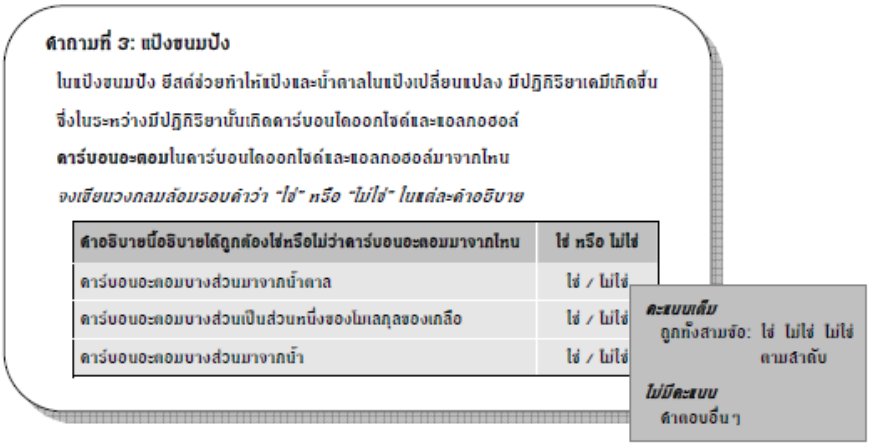
3.2 รูปแบบข้อสอบโครงการประเมินผลนักเรียนนานาชาติ

รูปแบบข้อสอบที่ใช้ในโครงการประเมินผลนักเรียนนานาชาติ ปี 2009 ประกอบด้วยลักษณะข้อสอบแบบต่างๆ ทั้งนี้ข้อสอบแบบเลือกตอบมีสัดส่วนประมาณหนึ่งในสามของข้อสอบทั้งหมด ซึ่งให้นักเรียนเลือกหนึ่งคำตอบจากสี่ตัวเลือก ส่วนข้อสอบแบบอื่นๆ นักเรียนต้องเขียน

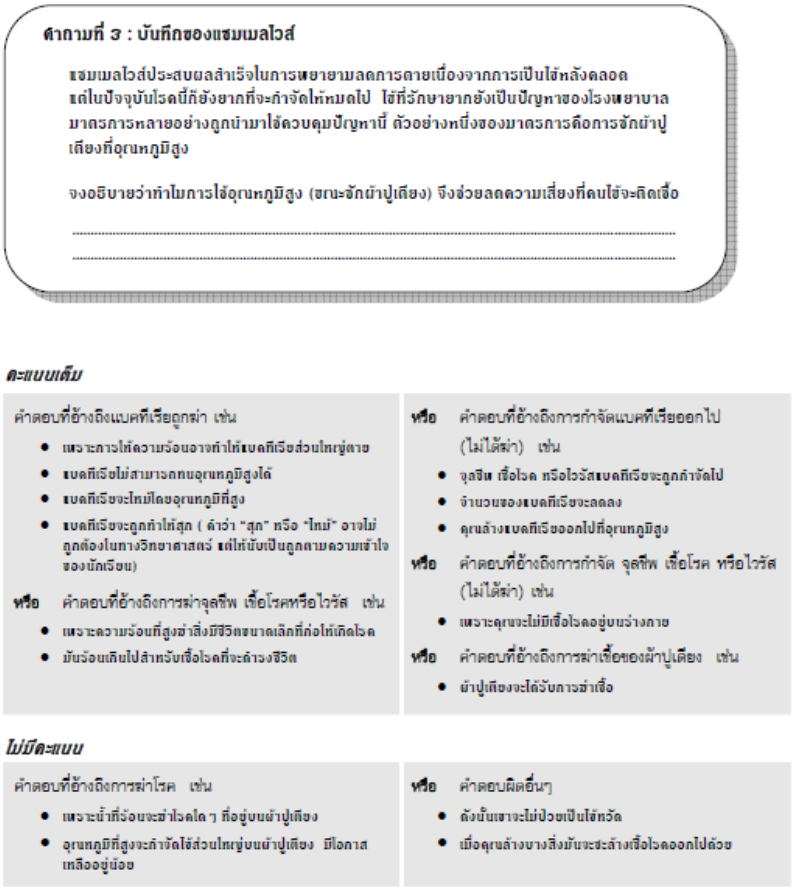
คำตอบ อาจเป็นการเติมคำ เขียนตอบสั้นๆ หรือแบบเลือกตอบเชิงซ้อน (complex multiple-choice) ซึ่งนักเรียนต้องเลือกใช้ตัวเลือกหลายตัวประกอบกัน นอกจากนี้ข้อสอบประเภทสร้างคำตอบแบบปิดและสร้างคำตอบแบบเปิดเป็นข้อสอบที่นักเรียนต้องใช้กระบวนการทางความคิดเชิงสติปัญญา (cognitive Process) ดังนั้น ข้อสอบจึงเปิดโอกาสให้นักเรียนสร้างคำอธิบายและสื่อสารให้ผู้อื่นเข้าใจความคิดและการใช้เหตุผลของนักเรียนในการตอบ โดยมีเกณฑ์การตรวจให้คะแนนสำหรับข้อสอบที่มีการตรวจให้คะแนนทวิวิภาค เช่น ข้อสอบประเภทเลือกตอบและข้อสอบแบบตอบอิสระถ้าตอบข้อสอบได้ถูกต้องจะได้คะแนนเต็ม 1 คะแนน ถ้าตอบผิดจะได้ 0 คะแนน ส่วนข้อสอบที่มีการตรวจให้คะแนนพหุวิภาค เช่น ข้อสอบเลือกตอบเชิงซ้อน และข้อสอบแบบเปิด ถ้าตอบข้อสอบได้ถูกต้องจะได้คะแนนเต็ม 2 คะแนน ถ้าตอบถูกบางส่วนจะได้ 1 คะแนน และถ้าตอบผิดจะได้ 0 คะแนน ตัวอย่างข้อสอบประเภทต่างๆ และการตรวจให้คะแนน แสดงได้ตามรูปภาพที่ 8-11



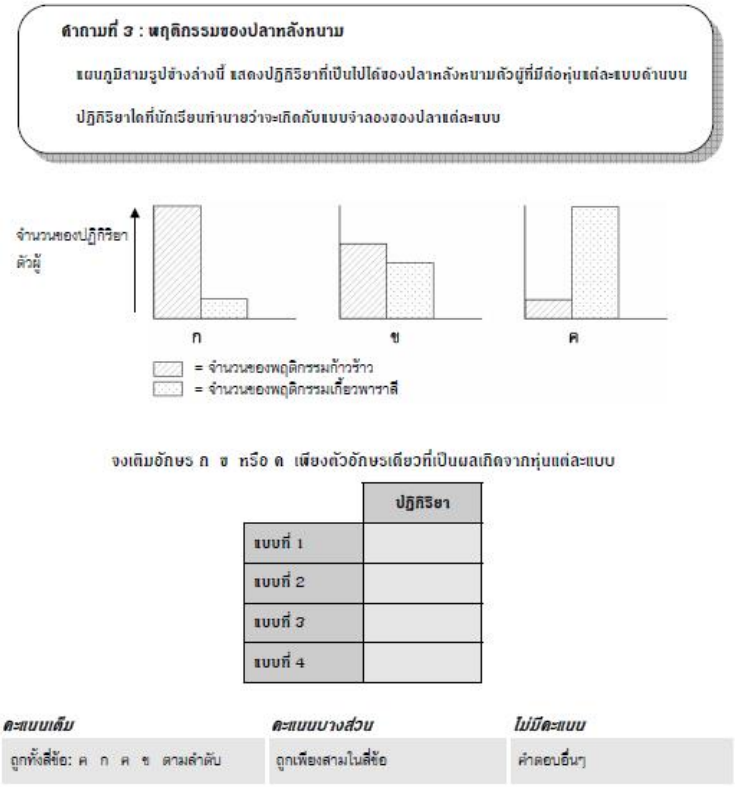
รูปภาพที่ 8 ตัวอย่างข้อสอบประเภทเลือกตอบและการให้คะแนน มาจากสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (2551)



รูปภาพที่ 9 ตัวอย่างข้อสอบเลือกตอบเชิงซ้อนและการให้คะแนน มาจากสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (2551)



รูปภาพที่ 10 ตัวอย่างข้อสอบแบบตอบอิสระแบบปิดและการให้คะแนน มาจากสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (2551)



รูปภาพที่ 11 ตัวอย่างข้อสอบแบบตอบอิสระแบบเปิดและการให้คะแนน มาจากสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (2551)

จุฬาลงกรณ์มหาวิทยาลัย

3.3 ข้อมูลสถานศึกษาจากการสำรวจในโครงการประเมินผลนักเรียนนานาชาติ

โครงการประเมินผลนักเรียนนานาชาติ (PISA) ปี 2009 ใช้เครื่องมือทั้งแบบสอบเพื่อวัดการรู้เรื่องทั้ง 3 ด้าน และแบบสอบถามเพื่อเก็บข้อมูลพื้นฐานทั้งของนักเรียนและสถานศึกษา โดยผู้ที่ให้ข้อมูลประกอบด้วยนักเรียนและผู้บริหารสถานศึกษา ทั้งนี้การให้ข้อมูลจะเกิดขึ้นภายหลังจากการตอบทำแบบทดสอบเสร็จเรียบร้อย ข้อมูลสถานศึกษาจากการสำรวจสามารถจำแนกออกเป็น 2 ระดับ คือ ระดับสถานศึกษา ได้แก่ โครงสร้างองค์กรภายในโรงเรียน นักเรียนและครู ทรัพยากรของโรงเรียน หลักสูตรและการประเมินผล สภาพบรรยากาศของโรงเรียน และนโยบายและการปฏิบัติของโรงเรียน และระดับนักเรียน ได้แก่ ภูมิหลังของนักเรียน ครอบครัว กิจกรรมการอ่านหนังสือ ระยะเวลาในการเรียนรู้ สภาพโรงเรียน ห้องสมุด และยุทธวิธีในการอ่านและสร้างความเข้าใจตำราเรียน รายละเอียดความหมายของข้อมูลที่ใช้ใน PISA 2009 จำแนกตามระดับข้อมูลรายละเอียดตามตารางที่ 11

ตารางที่ 11 สเกลการวัดและความหมายของตัวแปร

ตัวแปร		สเกลการวัดและความหมาย
โครงสร้างองค์กรภายในโรงเรียน (Structure and Organization of the school)	ระดับชั้นในโรงเรียน (grade level)	ตัวแปรจัดประเภท แบ่งเป็นชั้น ป.1 ถึง ม.6 และไม่มีระดับชั้น ให้ระบุ ใช่ หรือ ไม่ใช่
	ขนาดสถานศึกษา (school size)	ตัวแปรจัดประเภท แสดงจำนวนนักเรียนในสถานศึกษา จำแนกเป็น ขนาดใหญ่ (นักเรียนมากกว่า 1,500 คน), ขนาดปานกลาง (นักเรียน 500 – 1,500 คน), และขนาดเล็ก (นักเรียน 1– 499 คน)
	ประเภทสถานศึกษา (type of school)	ตัวแปรจัดประเภท ให้ระบุ โรงเรียนเอกชน หรือ โรงเรียนรัฐบาล
	ร้อยละงบประมาณของรัฐที่จัดสรรให้สถานศึกษา (percent of total funding)	ตัวแปรต่อเนื่อง แบ่งเป็น จากรัฐบาล ค่าธรรมเนียมต่างๆ การบริจาค หรืออื่น ให้ระบุเป็นเปอร์เซ็นต์ที่ได้ หรือระบุ 0 ถ้าไม่มีแหล่งเงินสนับสนุน
	สังกัดของสถานศึกษา	ตัวแปรจัดประเภท (เป็นตัวแปรที่เก็บเพิ่มจากหน่วยงาน สสวท.) จำแนกเป็นสถานศึกษา 8 สังกัด ได้แก่ สพฐ.1, สพฐ.2, สช., สกอ., กศท., กทม., อศ.1, และ อศ.2
	ที่ตั้งสถานศึกษา (community)	ตัวแปรจัดประเภท แสดงขนาดของชุมชนที่สถานศึกษาตั้งอยู่ จำแนกเป็น หมู่บ้าน (ประชาชนน้อยกว่า 3,000 คน), เมืองเล็ก (ประชาชน 3,000-ประมาณ 15,000 คน), เมือง (ประชาชน 15,000 -100,000 คน), เมืองใหญ่ (ประชาชน 100,000 -1,000,000 คน), และเมืองใหญ่มาก (ประชาชนมากกว่า 1,000,000 คน)
	ความสนใจของผู้ปกครองในการเลือกเรียน (choosing a school)	ตัวแปรจัดประเภท ให้ระบุ มีโรงเรียน มากกว่า 2 แหล่งในพื้นที่, มีโรงเรียน 1 แหล่งในพื้นที่, หรือ ไม่มีโรงเรียนในพื้นที่
	จำนวนโรงเรียนในพื้นที่ที่นักเรียนสามารถเข้าเรียนได้ (area)	ตัวแปรต่อเนื่อง จำแนกเป็น จำนวนโรงเรียนเอกชนหรือโรงเรียนรัฐบาล ให้ระบุจำนวนโรงเรียน

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 11 สเกลการวัดและความหมายของตัวแปร (ต่อ)

ตัวแปร		สเกลการวัดและความหมาย
นักเรียนและครู (The student body and teachers)	จำนวนนักเรียนที่เข้าเรียนในแต่ละปี (school enrollment)	ตัวแปรต่อเนื่อง จำแนกเป็น นักเรียนชายและนักเรียนหญิง ให้ระบุจำนวนนักเรียน
	เปอร์เซ็นต์นักเรียนที่เรียนซ้ำชั้น (school repeated)	ตัวแปรต่อเนื่อง จำแนกเป็น ระดับชั้น ม.ปลาย และ ม.ต้น ให้ระบุเปอร์เซ็นต์ซ้ำชั้น
	วุฒิการศึกษาของครู (teacher quality)	ตัวแปรจัดประเภท จำแนกเป็น ครูทั้งหมด, ครูที่มีคุณวุฒิตรงสาขา, และครูที่มีวุฒิการศึกษาน้อยปริญญาตรี ให้ระบุจำนวนแยกเป็นทำงานเต็มเวลาและทำงานเป็นช่วงเวลา
	สัดส่วนนักเรียนต่อครู (student-teacher ratio)	ตัวแปรต่อเนื่อง เป็นค่าที่แสดงสัดส่วนนักเรียนต่อครูของสถานศึกษา
ทรัพยากรของโรงเรียน (The school's resources)	อัตราส่วนคอมพิวเตอร์ต่อนักเรียน (student-computer ratio)	ตัวแปรต่อเนื่อง จำแนกเป็น จำนวนนักเรียน ม. 4, จำนวนคอมพิวเตอร์สำหรับนักเรียน, จำนวนคอมพิวเตอร์ที่เชื่อมต่ออินเทอร์เน็ต, และจำนวนคอมพิวเตอร์ที่ใช้ได้จริงสำหรับนักเรียนทุกคน ให้ระบุจำนวน
	ความสามารถของโรงเรียนต่อการจัดสรรทรัพยากรทางการศึกษา (school's capacity)	ตัวแปรจัดประเภท จำแนกเป็น 13 รายการ ให้ระบุไม่ทั้งหมด น้อยมาก มีบ้าง หรือ มาก
	ทรัพยากรการเรียน (the school's educational resources)	ตัวแปรต่อเนื่อง เป็นค่าดัชนีชี้วัดที่ได้มาจาก ข้อรายการพื้นฐาน 7 ประการที่ใช้วัดองค์ประกอบตามศักยภาพด้านอุปกรณ์การเรียนการสอนของโรงเรียน ได้แก่ ห้องเล็บวิทยาศาสตร์ คู่มือการเรียนการสอน คอมพิวเตอร์ อินเตอร์เน็ต ซอฟแวร์ที่ใช้ในการเรียนการสอน วัสดุอุปกรณ์ในห้องสมุด และโสตทัศนวัสดุ ทั้งนี้ทรัพยากรการเรียนมีค่าเฉลี่ยเป็นศูนย์ ถ้าค่าดัชนีเป็นบวกหมายความว่าทรัพยากรการเรียนมีความเพียงพอหรือไม่ขาดแคลน และถ้าดัชนีเป็นลบหมายความว่าทรัพยากรการเรียนขาดแคลน

ตารางที่ 11 สเกลการวัดและความหมายของตัวแปร (ต่อ)

ตัวแปร		สเกลการวัดและความหมาย
หลักสูตรและการประเมินผลของสถานศึกษา (School curriculum and assessment)	นโยบายของโรงเรียน (school's policy)	ตัวแปรจัดประเภท จำแนกเป็น นักเรียนถูกแบ่งกลุ่มความสามารถในชั้นที่ต่างกัน และนักเรียนถูกแบ่งกลุ่มความสามารถในชั้นเดียวกัน ให้ระบุ ทุกวิชา บางวิชา หรือ ไม่เกิดขึ้นในวิชาใดเลย
	กิจกรรมสร้างสรรค์ต่อนักเรียน (activities)	ตัวแปรจัดประเภท จำแนกเป็น 14 รายการ ให้ระบุ ใช่ หรือไม่ใช่
	การประเมินผล (assessment)	ตัวแปรจัดประเภท จำแนกเป็น การทดสอบมาตรฐาน, แบบสอบที่สร้างจากครู, การให้คะแนนอย่างยุติธรรมของครู, แฟ้มผลงานของนักเรียนม และงานที่รับมอบหมาย/โครงงาน/การบ้าน ให้ระบุ ไม่เคย 1-2 ครั้งต่อปี 3-5 ครั้งต่อปี ทุกเดือน หรือมากกว่าหนึ่งครั้งต่อเดือน
	วัตถุประสงค์ในการประเมินผล (assessed purposes)	ตัวแปรจัดประเภท 8 รายการ ให้ระบุ ใช่ หรือไม่ใช่
	หลักสูตรการวัดและประเมินผล (curriculum and assessment)	ตัวแปรต่อเนื่อง เป็นค่าดัชนีที่ได้มาจากคะแนนรวมรายงานผู้ที่มีความรับผิดชอบด้านหลักสูตรและการวัดและประเมินผลของโรงเรียน ได้แก่ นโยบายการประเมินผลนักเรียน คู่มือที่ใช้เนื้อหาในหลักสูตร และหลักสูตรที่สนับสนุน ทั้งนี้หลักสูตรการวัดและประเมินภายในสถานศึกษามีค่าเฉลี่ยเป็นศูนย์ ถ้าค่าดัชนีเป็นบวกหมายความว่ามีหลักสูตรและการประเมินผลเหมาะสม และถ้าดัชนีเป็นลบหมายความว่าหลักสูตรและการประเมินผลเกิดขึ้นไม่ครบถ้วน
สภาพบรรยากาศของโรงเรียน (School climate)	การเรียนรู้ของนักเรียน (learning of student)	ตัวแปรจัดประเภท 13 รายการ ให้ระบุ ไม่เคยเกิดขึ้น เกิดขึ้นน้อย เกิดขึ้นบางครั้ง หรือเกิดขึ้นมาก
	ความคาดหวังของผู้ปกครอง (expectation)	ตัวแปรจัดประเภท ให้ระบุ อย่างสม่ำเสมอ จากผู้ปกครอง จากผู้ปกครองส่วนใหญ่ หรือไม่ค่อยให้ความสนใจ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 11 สเกลการวัดและความหมายของตัวแปร (ต่อ)

ตัวแปร		สเกลการวัดและความหมาย
นโยบายและการปฏิบัติของโรงเรียน (School policies and practices)	ปัจจัยในการพิจารณาเข้าเรียนที่โรงเรียน (factors)	ตัวแปรจัดประเภท 7 รายการ ระบุ ไม่เคย บางเวลา หรือ ทุกครั้ง
	สาเหตุการย้ายโรงเรียน (transfer to another school)	ตัวแปรจัดประเภท 6 รายการ ระบุ ไม่มีแนวโน้ม มีแนวโน้ม หรือมีแนวโน้มมาก
	ความรับผิดชอบของโรงเรียนต่อผู้ปกครอง (school's accountability to parents)	ตัวแปรจัดประเภท จำแนกเป็น 3 รายการ ระบุ ใช่ หรือ ไม่ใช่
	ข้อมูลผลสัมฤทธิ์ของโรงเรียน (achievement data)	ตัวแปรจัดประเภท จำแนกเป็น 5 รายการ ระบุ ใช่ หรือ ไม่ใช่
	การรับผิดชอบติดตามภาระงาน (responsibility)	ตัวแปรจัดประเภท จำแนกเป็น 12 รายการ ระบุ ผู้อำนวยการ ครู คณะกรรมการระดับโรงเรียน ตัวแทนจากรัฐหรือท้องถิ่น หรือกระทรวงศึกษาธิการ
	ระดับการขาดแคลนครู (shortage of science teacher)	ตัวแปรจัดประเภท แสดงระดับการขาดแคลนครูวิทยาศาสตร์ จำแนกเป็น 4 ระดับคือ ไม่ขาดแคลน ขาดแคลนน้อย ขาดแคลนเป็นบางส่วน ขาดแคลนจำนวนมาก
นโยบายและการปฏิบัติของโรงเรียน (School policies and practices)	การตัดสินใจเกี่ยวกับคณะทำงานภายในโรงเรียน (decision making about staffing)	ตัวแปรจัดประเภท จำแนกเป็น 7 รายการ ให้ระบุ เจ้าหน้าที่ งบประมาณ เนื้อหาสื่อการเรียนการสอน การประเมินการปฏิบัติงาน
	การบริหารจัดการภายในโรงเรียน (management)	ตัวแปรจัดประเภท จำแนกเป็น 14 รายการ ให้ระบุ ไม่เคย ไม่ค่อย บางครั้ง บ่อยครั้ง
	เปอร์เซ็นต์นักเรียนที่มีสิทธิ์ได้รับประทานอาหารกลางวัน (percentage of students who were eligible of free- or reduced-price lunches)	ตัวแปรต่อเนื่อง ระบุเปอร์เซ็นต์ของนักเรียนที่เหมาะสม

ตารางที่ 11 สเกลการวัดและความหมายของตัวแปร (ต่อ)

ตัวแปร		สเกลการวัดและความหมาย
ตัวแปรระดับนักเรียน ภูมิหลังของนักเรียน (background)	ระดับชั้น (grade)	ตัวแปรจัดประเภท ให้ระบุระดับชั้นที่ศึกษา
	ระยะเวลาที่เรียน (How long have you been in this school)	ตัวแปรจัดประเภท จำแนกเป็น น้อยกว่า 1ปี 1-2 ปี 3-4 ปี และมากกว่า 4 ปี
	เพศ (sex)	ตัวแปรจัดประเภท ให้ระบุ ชาย หรือ หญิง
	การเข้าเรียนก่อนชั้นอนุบาล (attend pre-school)	ตัวแปรจัดประเภท ให้ระบุ ไม่เคย เคย 1 ปี หรือน้อยกว่า หรือเคยมากกว่า 1 ปี
	การเข้าเรียนระดับชั้นอนุบาล (attend kindergarten)	ตัวแปรจัดประเภท ระบุ เคยเข้าเรียน หรือไม่เคยเข้าเรียน
	อายุเมื่อเริ่มเรียนชั้น ป.1 (started first grade)	ตัวแประต่อเนื่อง ระบุอายุที่เข้าเรียน
	การเรียนซ้ำชั้น (repeated a grade)	ตัวแปรจัดประเภท แยกเป็นระดับชั้นอนุบาล ป.1-ป.6 ม.1-ม.3 และ ม.4-ม.6 ให้ระบุ ไม่เคย หนึ่งครั้ง หรือมากกว่า 2 ครั้ง
	ระดับการศึกษาสูงสุดที่คาดหวัง (level of school you expect)	ตัวแปรจัดประเภท ให้ระบุ น้อยกว่ามัธยมศึกษา มัธยมศึกษา ประกาศนียบัตรอนุปริญญา(2 ปี) ปริญญาตรี(4 ปี) ปริญญาโท หรือ ปริญญาเอก/วิชาชีพเฉพาะทางเช่น หมอ ทนาย
ครอบครัว (Your family and your home)	การอาศัยร่วมกับสมาชิกภายในครอบครัว (live at home)	ตัวแปรจัดประเภท ให้ระบุ แม่(รวมแม่บุญธรรม) พ่อ(รวมพ่อบุญธรรม) พี่ชาย/น้องชาย (รวมพี่น้องบุญธรรม) น้องสาว/พี่สาว(รวมพี่น้องบุญธรรม) หลานๆ หรืออื่นๆเช่น ญาติ
	ระดับการศึกษาสูงสุดของมารดา (highest level of schooling by your mother)	ตัวแปรจัดประเภท ให้ระบุ จบ ม.6 จบ ม.3 จบ ป.6 หรือ ไม่จบ/จบ ป.6

ตารางที่ 11 สเกลการวัดและความหมายของตัวแปร (ต่อ)

ตัวแปร		สเกลการวัดและความหมาย
	ปริญญาหรือใบประกาศนียบัตรที่มารดาถือครอง (certificates or diplomas)	ตัวแปรจัดประเภท จำแนกเป็น ปริญญาโท ปริญญาเอก หรืออาชีพเฉพาะ เช่น หมอ หรือ ทนาย ปริญญาตรี น้อยกว่าปริญญาตรี ประกาศนียบัตรประเภทอาชีวะ ให้ระบุ ใช่หรือไม่ใช่
	การทำงานปัจจุบันของมารดา (currently doing)	ตัวแปรจัดประเภท ให้ระบุ ทำงานเต็มเวลา ทำงานเป็นช่วงเวลา ไม่ทำงานหรือกำลังหางานอื่นๆ เช่น ทำงานบ้านหรือเกษียณอายุ
	ระดับการศึกษาสูงสุดของบิดา (highest level of schooling by your father)	ตัวแปรจัดประเภท ให้ระบุ จบ ม.6 จบม.3 จบ ป.6 หรือ ไม่จบ จบ ป.6
	ปริญญาหรือใบประกาศนียบัตรที่บิดาถือครอง (certificates or diplomas)	ตัวแปรจัดประเภท จำแนกเป็น ปริญญาโท เอก หรืออาชีพเฉพาะ เช่น หมอ หรือ ทนาย ปริญญาตรี น้อยกว่าปริญญาตรี ประกาศนียบัตรประเภทอาชีวะ ให้ระบุ ใช่หรือไม่ใช่
	การทำงานปัจจุบันของบิดา (currently doing)	ตัวแปรจัดประเภท ให้ระบุ ทำงานเต็มเวลา ทำงานเป็นช่วงเวลา ไม่ทำงานหรือกำลังหางานอื่นๆ เช่น ทำงานบ้านหรือเกษียณอายุ
	สิ่งของภายในบ้าน (your home)	ตัวแปรจัดประเภท แบ่งเป็น 17 รายการ ให้ระบุ ใช่ หรือ ไม่ใช่
	จำนวนสิ่งของที่มีในบ้าน (how many of these are there at your home)	ตัวแปรจัดประเภท แบ่งเป็น 5 รายการ ให้ระบุ ไม่มี มีหนึ่ง สอง สามหรือมากกว่า
	จำนวนหนังสือที่มีในบ้าน (books)	ตัวแปรจัดประเภท ให้ระบุ จำนวนหนังสือ 6 รายการ
	เศรษฐานะของครอบครัว (economic, social and cultural status)	ตัวแปรต่อเนื่อง เป็นดัชนีชี้วัดที่ได้มาจากจากคะแนนรวมที่วัดได้จากสถานะทางการงานการศึกษาของพ่อแม่ ฐานะทางเศรษฐกิจทรัพยากรด้านการศึกษาและวัฒนธรรมของครอบครัวของเด็กนักเรียน โดยที่เศรษฐานะของครอบครัวมีค่าเฉลี่ยเป็นศูนย์ ถ้าค่าดัชนีเป็นบวกหมายความว่าเศรษฐานะของครอบครัวดี ถ้าดัชนีเป็นลบหมายความว่าเศรษฐานะของครอบครัวไม่ดี

ตารางที่ 11 สเกลการวัดและความหมายของตัวแปร (ต่อ)

ตัวแปร		สเกลการวัดและความหมาย
	แหล่งทรัพยากรการศึกษาที่บ้าน (home educational resources)	ตัวแปรต่อเนื่อง ดัชนีชี้วัดที่ได้มาจากคะแนนรวมที่วัดได้จากรายการสิ่งของที่มีอยู่ภายในบ้าน เช่น พจนานุกรม คอมพิวเตอร์ โต๊ะทำงาน หนังสือเรียน โดยที่แหล่งทรัยากรการศึกษาที่บ้านมีค่าเฉลี่ยเป็นศูนย์ ถ้าค่าดัชนีเป็นบวกหมายความว่ามีแหล่งทรัพยากรการศึกษาที่บ้านดี ถ้าดัชนีเป็นลบหมายความว่ามีแหล่งทรัพยากรการศึกษาที่บ้านไม่ดี
	ความมั่งคั่งของครอบครัว (family wealth)	ตัวแปรต่อเนื่อง เป็นค่าดัชนีชี้วัดที่ได้มาจากคะแนนรวมที่วัดได้จากสิ่งของเครื่องใช้ต่างๆ ภายในบ้าน เช่น อินเทอร์เน็ต เครื่องล้างจาน เครื่องเล่นดีวีดี โทรศัพท์เคลื่อนที่ โทรทัศน์ รถ และห้องน้ำในบ้านที่มีฝักบัวหรืออ่างอาบน้ำ โดยที่ความมั่งคั่งของครอบครัวมีค่าเฉลี่ยเป็นศูนย์ ถ้าดัชนีเป็นบวกหมายความว่าครอบครัวมีความมั่งคั่ง ถ้าดัชนีเป็นลบหมายความว่าครอบครัวไม่มีความมั่งคั่ง
	แหล่งทรัพยากรทางไอซีที (ICT resources at home)	ตัวแปรต่อเนื่อง เป็นค่าดัชนีชี้วัดที่ได้มาจากคะแนนรวมแหล่งทรัพยากรทางไอซีทีของนักเรียน โดยที่แหล่งทรัพยากรทางไอซีทีมีค่าเฉลี่ยเป็นศูนย์ ถ้าดัชนีเป็นบวกหมายความว่ามีแหล่งทรัพยากรทางไอซีทีเพียงพอ ถ้าดัชนีเป็นลบหมายความว่ามีแหล่งทรัพยากรทางไอซีทีขาดแคลน
กิจกรรมการอ่านหนังสือ (your reading activities)	จำนวนเวลาที่ใช้ในการอ่าน (spend reading)	ตัวแปรจัดประเภท ให้ระบุระยะเวลาที่ใช้ในการอ่าน 5 รายการ
	ทัศคติเกี่ยวกับการอ่านหนังสือ (attitude)	ตัวแปรเรียงลำดับ แบ่งเป็น 11 รายการ ให้ระบุ ไม่เห็นด้วยอย่างยิ่ง ไม่เห็นด้วย เห็นด้วย และเห็นด้วยอย่างยิ่ง
	ประเภทหนังสือที่อ่าน (materials)	ตัวแปรจัดประเภท แบ่งเป็น 5 รายการ เลือก ความถี่ 5 ช่วง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 11 สเกลการวัดและความหมายของตัวแปร (ต่อ)

ตัวแปร		สเกลการวัดและความหมาย
	กิจกรรมการอ่าน (reading activities)	ตัวแปรจัดประเภท แบ่งเป็น 8 รายการ เลือกจำนวนครั้งที่ทำ
	ยุทธวิธีในการจดจำ (memorize)	ตัวแปรจัดประเภท แบ่งเป็น 13 รายการ ให้ระบุความถี่ เกือบไม่เคย บางครั้ง บ่อยครั้ง หรือ ทุกครั้ง
ระยะเวลาในการเรียนรู้ (Learning time)	จำนวนนาทีในการเรียนวิชาวิทยาศาสตร์ (class period)	ตัวแปรต่อเนื่อง ให้ระบุจำนวนนาทีในการเรียนวิทยาศาสตร์
	การใช้เวลาเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา (out-of-time lessons)	ตัวแปรจัดประเภท ให้ระบุ ไม่เคยเรียนพิเศษ, น้อยกว่า 2 ชั่วโมงต่อสัปดาห์, 2 ชั่วโมงแต่ไม่เกิน 4 ชั่วโมงต่อสัปดาห์, 4 ชั่วโมงแต่ไม่เกิน 6 ชั่วโมงต่อสัปดาห์, หรือ 6 ชั่วโมงหรือมากกว่าต่อสัปดาห์
สภาพโรงเรียน (Your school)	การเรียนรู้ในโรงเรียน (learned in school)	ตัวแปรจัดลำดับ แบ่งเป็น 4 กลุ่ม ให้ระบุ ไม่เห็นด้วยอย่างยิ่ง ไม่เห็นด้วย เห็นด้วย เห็นด้วยอย่างยิ่ง
	ทัศนคติต่อครู (teacher at you school)	ตัวแปรจัดลำดับ แบ่งเป็น 5 กลุ่ม ให้ระบุ ไม่เห็นด้วยอย่างยิ่ง ไม่เห็นด้วย เห็นด้วย เห็นด้วยอย่างยิ่ง
ห้องสมุด (Libraries)	กิจกรรมที่เกิดขึ้น (activities)	ตัวแปรจัดประเภท แบ่งเป็น 7 ประเภท ให้ระบุ ไม่เคยเกิดขึ้น 2 ครั้งต่อปี 1 ครั้งต่อเดือน หลายครั้งต่อเดือน หรือ หลายครั้งต่อสัปดาห์
	จำนวนห้องสมุดในโรงเรียน (school library)	ตัวแปรจัดประเภท ให้ระบุ มี หรือ ไม่มี
ยุทธวิธีในการอ่านและสร้างความเข้าใจตำราเรียน (Your strategies in reading and understanding texts)	ความเข้าใจและจดจำในข้อมูลตำราเรียน (understand and remember)	ตัวแปรเรียงลำดับ แบ่งเป็น 6 ยุทธวิธี ให้ระบุ 6 สเกล ตั้งแต่ ไม่มีประโยชน์ ไปจนถึง มีประโยชน์อย่างมาก

3.4 งานวิจัยที่เกี่ยวข้องโครงการประเมินผลนักเรียนนานาชาติ

จากการศึกษางานวิจัยที่เกี่ยวข้องกับโครงการประเมินผลนักเรียนนานาชาติ พบว่า มีการศึกษาที่เกี่ยวข้องกับโครงการหลายด้าน จำแนกออกเป็น 4 กลุ่ม ได้แก่ กลุ่มที่ 1 ศึกษาวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบโดยใช้สถิติในการวิเคราะห์ (Le, 2009; Oliveri et al., 2012; Kankaras & Moors, 2014) กลุ่มที่ 2 ศึกษาวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบโดยใช้สถิติและผู้เชี่ยวชาญในการวิเคราะห์ (Liu & Wilson, 2009; Yildirim & Berberoglu, 2009) กลุ่มที่ 3 ศึกษาวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบโดยเน้นที่การศึกษาตัวแปร (Yildirim & Yildirim, 2011) และกลุ่มที่ 4 ศึกษาความสัมพันธ์ระหว่างตัวแปร (Van de gaer et al., 2012)

กลุ่มที่ศึกษาวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบโดยใช้สถิติในการวิเคราะห์ ได้แก่ Le (2009) ได้ทำการศึกษาการทำหน้าที่ต่างกันของข้อสอบระหว่างเพศข้ามประเทศ จากแบบทดสอบวิชาวิทยาศาสตร์ ในการศึกษาครั้งนี้โดยแบ่งเป็น 4 มิติได้แก่ จุดเน้น เนื้อหา ความสามารถ และความรู้ทางวิทยาศาสตร์ โดยเก็บข้อมูลจากกลุ่มแบบทดสอบ 60 ชุดภาษา จากประเทศที่เข้าร่วม 50 ประเทศ จากนักเรียนอายุ 15 ปีจำนวน 83,000 คน โดยใช้การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบตามแนวคิด IRT พบว่า เพศมีมิติที่แตกต่างกันทางด้านข้อสอบ 5 ลักษณะ ส่วนข้อสอบประเภทหลายตัวเลือกและข้อสอบประเภทให้คำตอบแบบปิดมีแนวโน้มจะเข้าข้างนักเรียนชายมากกว่านักเรียนหญิง

นอกจาก Le (2009) แล้ว Oliveri และคณะ (2012) ศึกษาวิธีวิทยาในการสำรวจความเท่าเทียมในการวัดระดับข้อสอบและแบบสอบในการประเมินข้อสอบระดับนานาชาติ ในการศึกษาครั้งนี้เป็นการวัดการแก้ปัญหาในโครงการประเมินผลนักเรียนนานาชาติ (PISA) ปี 2003 ของกลุ่มนักเรียนที่ใช้ภาษาแคนาดาและภาษาฝรั่งเศส โดยอาศัยวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเปรียบเทียบกัน 3 วิธี ได้แก่ ทฤษฎีตอบสนองข้อสอบแบบพาราเมทริก ทฤษฎีตอบสนองข้อสอบแบบนันพาราเมทริกและการวัดการทดถอยโลจิสติกเชิงอันดับ (ordinal logistic regression) พบว่า การทำหน้าที่ต่างกันของแบบสอบเข้าข้างนักเรียนที่พูดภาษาอังกฤษและผลการทดสอบความไม่เท่าเทียมกันในระดับข้อสอบไม่สามารถตรวจสอบได้ในระดับแบบสอบ

ส่วน Kankaras และ Moors (2014) ได้วิเคราะห์และเปรียบเทียบข้ามวัฒนธรรมของคะแนนโครงการประเมินผลนักเรียนนานาชาติ ในปี 2009 ซึ่งเป็นวัดระดับความเท่าเทียมกันโดยใช้วิธีตอบสนองข้อสอบ (IRT) และวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) พบว่า ผลจากการวัดความไม่เท่าเทียมกันเกิดขึ้นจากข้อคำถามในแบบสอบทั้งสามฉบับ ได้แก่ คณิตศาสตร์ วิทยาศาสตร์ และการอ่าน โดยมีขนาดระดับปานกลางซึ่งพิจารณาจากค่าเฉลี่ย และเมื่อทดสอบ

การทำหน้าที่ต่างกันแบบเอกรูป (uniform DIF) พบว่า ส่วนใหญ่เอเชียตะวันออกเฉียงใต้มีสเกลคะแนนสูงในช่วงแรกและมีลักษณะโมเดลเป็นเอกพันธ์ (homogeneous model)

นอกจากนี้กลุ่มที่ศึกษาวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบโดยใช้สถิติและผู้เชี่ยวชาญในการวิเคราะห์ ได้แก่ Liu และ Wilson (2009) ได้ศึกษาความแตกต่างระหว่างเพศในการประเมินวิชาคณิตศาสตร์ของโครงการประเมินผลนักเรียนนานาชาติ (PISA) ตั้งแต่ปี 2000 ถึง 2003 โดยใช้โมเดลราสช์พหุมิติ (a multidimensional rasch model) ในการประมาณค่าความสามารถนักเรียน พบว่า ข้อสอบประเภทหลายตัวเลือกเข้าข้างกลุ่มนักเรียนชายมากกว่านักเรียนหญิง ความแตกต่างระหว่างข้อสอบทั้งขนาดข้อสอบ มิติในการออกข้อสอบและรูปแบบข้อสอบ เป็นผลให้เกิดความแตกต่างในการทำข้อสอบทั้งกลุ่มนักเรียนชายและนักเรียนหญิง

นอกจาก Liu และ Wilson (2009) ยังมี Yildirim และ Berberoglu (2009) ศึกษาการทำหน้าที่ต่างกันของข้อสอบเชิงสถิติและโดยผู้ตัดสิน โดยการวิเคราะห์ข้อสอบการรู้เรื่องคณิตศาสตร์ของโครงการประเมินผลนักเรียนนานาชาติ (PISA) ปี 2003 ซึ่งใช้สถิติในการวิเคราะห์ได้แก่ restricted factor analysis (RFA), mantel-haenzsel (MH) และ response theory likelihood ratio analysis (IRT-LR) ส่วนกลุ่มผู้ตัดสินรายงานว่า ทักษะเชิงพุทธิพิสัยวัดได้โดยลักษณะข้อสอบ ความคลาดเคลื่อนในการตีความและการใช้ปริมาณของคำ ซึ่งทั้งสามแหล่งเป็นแหล่งความคลาดเคลื่อนที่อาจจะเป็นสาเหตุของการเกิดการทำหน้าที่ต่างกันของข้อสอบ และผลการศึกษาเสนอแนะว่าการวิเคราะห์เนื้อหาเบื้องต้นของข้อสอบอย่างเป็นระบบอาจจะช่วยป้องกันการเกิดการทำหน้าที่ต่างกันของข้อสอบ

กลุ่มที่ศึกษาวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบโดยเน้นที่การศึกษาตัวแปร ได้แก่ Yildirim และ Yildirim (2011) ศึกษาความสัมพันธ์ของชุมชนในฐานะที่เป็นตัวแปรที่สอดคล้อง (Matching variables) ในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ โดยใช้ข้อมูลวิชาคณิตศาสตร์จากโครงการประเมินผลนักเรียนนานาชาติ พบว่า นักเรียนที่ใช้เวลาในการเรียนรู้นอกห้องเรียนสามารถตรวจสอบได้ว่าเป็นตัวแปรสอดคล้องที่เพิ่มขึ้นมา และผลการทำหน้าที่ต่างกันของข้อสอบแบบหลายตัวแปรแสดงให้เห็นว่าตัวแปรดังกล่าวอาจจะมีความสัมพันธ์กับแหล่งของการเกิดการทำหน้าที่ต่างกันของข้อสอบในบางข้อสอบ

ส่วน กลุ่มที่ศึกษาความสัมพันธ์ระหว่างตัวแปร ได้แก่ Van de gaer และคณะ (2012) ศึกษาแบบข้ามวัฒนธรรมของอิทธิพลกลุ่มเปรียบเทียบโดยการสำรวจความสัมพันธ์แบบไม่คงที่ระหว่างผลสัมฤทธิ์ทางการเรียนและความมั่นใจในตนเอง (self-confidence) พบว่า การศึกษาจากสมการถดถอยพหุระดับ (multilevel regression analyses) จากข้อมูลโครงการประเมินผลนักเรียนนานาชาติ ปี 2006 ภายในโรงเรียน คะแนนผลสัมฤทธิ์วิชาวิทยาศาสตร์ ความเชื่อมั่นในตนเองวิชา

วิทยาศาสตร์ มีความสัมพันธ์ในทิศทางบวก ในขณะที่ระหว่างโรงเรียนและระหว่างประเทศมีความสัมพันธ์ในทิศทางลบ

จากการศึกษางานวิจัยที่เกี่ยวข้องกับโครงการประเมินผลนักเรียนนานาชาติ สามารถสรุปผลการวิเคราะห์องค์ความรู้ที่ได้รายละเอียดตามตารางที่ 12 ดังนี้

ตารางที่ 12 สรุปผลองค์ความรู้จากงานวิจัยโครงการประเมินผลนักเรียนนานาชาติ

ผู้วิจัย	วิธีการดำเนินงาน	องค์ความรู้ในการศึกษา
Le (2009)	ศึกษาการทำหน้าที่ต่างกันของข้อสอบระหว่างเพศข้ามประเทศ	วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบใช้วิธีตอบสนองข้อสอบ (IRT) ในการตรวจสอบความแตกต่างระหว่างเพศที่มีลักษณะแบบเอกรูป (uniform DIF)
Liu และ Wilson (2009)	ศึกษาความแตกต่างระหว่างเพศในการประเมินวิชาคณิตศาสตร์	ใช้โมเดลราสช์พหุมิติ (A multidimensional Rasch model) ในการประมาณค่าความสามารถนักเรียน
Kankaras และ Moors (2014)	วิเคราะห์และเปรียบเทียบข้ามวัฒนธรรมของคะแนน PISA	วัดระดับความเท่าเทียมกันโดยใช้วิธีตอบสนองข้อสอบ (IRT) และวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF)
Yildirim และ Berberoglu (2009)	ศึกษาการทำหน้าที่ต่างกันของข้อสอบเชิงสถิติและผู้ตัดสิน	สถิติในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบได้แก่ restricted factor analysis (RFA), mantel-haenzsel (MH) และ item response theory likelihood ratio analysis (IRT-LR) และตัดสินโดยผู้เชี่ยวชาญ
Yildirim และ Yildirim (2011)	ศึกษาความสัมพันธ์ของชุมชนในฐานะที่เป็นตัวแปรที่สอดคล้อง (matching variables) ในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ	ใช้ชุมชนเป็นตัวแปรที่สอดคล้อง (matching variables) ในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ

ตารางที่ 12 สรุปผลองค์ความรู้จากงานวิจัยโครงการประเมินผลนักเรียนนานาชาติ (ต่อ)

ผู้วิจัย	วิธีการดำเนินงาน	องค์ความรู้ในการศึกษา
Oliveri และคณะ (2012)	สำรวจความเท่าเทียมในการวัดระดับข้อสอบและแบบสอบ	การวัดการทำหน้าที่ต่างกันของข้อสอเปรียบเทียบกัน 3 วิธี ได้แก่ ทฤษฎีตอบสนองข้อสอบแบบพาราเมทริก ทฤษฎีตอบสนองข้อสอบแบบนันพาราเมทริก และการวัดการถดถอยโลจิสติกเชิงอันดับ (ordinal logistic regression)
Van de gaer และคณะ (2012)	ศึกษาแบบข้ามวัฒนธรรมของอิทธิพลกลุ่มเปรียบเทียบโดยการสำรวจความสัมพันธ์แบบไม่คงที่ระหว่างผลสัมฤทธิ์ทางการเรียนและความมั่นใจในตนเอง (self-confidence)	ใช้การวิเคราะห์ถดถอยพหุระดับ (multilevel regression analyses)

โครงการประเมินผลนักเรียนนานาชาติ หรือเรียกว่า โครงการ PISA (programme for international student assessment: PISA) เป็นโครงการศึกษาสำรวจความรู้และทักษะของนักเรียนกลุ่มอายุ 15 ปี ในประเทศสมาชิกองค์กรเพื่อความร่วมมือและพัฒนาทางเศรษฐกิจ (OECD) และประเทศร่วมโครงการที่ไม่ใช้สมาชิก โดยวัตถุประสงค์หลักของโครงการดังกล่าวเพื่อให้ข้อมูลกับประเทศสมาชิกว่าได้เตรียมความพร้อมของประชาชนให้มีศักยภาพสำหรับการแข่งขันในประชาชมโลกเพียงพอหรือไม่อย่างไรและเพื่อหาตัวชี้วัดคุณภาพการศึกษาในระดับนโยบาย โดยตัวชี้วัดที่สำคัญในการประเมินได้แก่ การรู้เรื่องคณิตศาสตร์ (mathematics Literacy) การรู้เรื่องการอ่าน (reading Literacy) และ การรู้เรื่องวิทยาศาสตร์ (scientific Literacy) การศึกษาของประเทศในโครงการ PISA จึงมีการประเมินผลทุกๆ สามปี แต่การประเมินผลแต่ละครั้งจะให้น้ำหนักสำหรับกับการประเมินต่างกัน โดยการประเมินผลถูกกำหนดเป็นรอบ (cycle) รอบแรก เริ่มต้นมาตั้งแต่ ปี 2000 จึงครอบรอบหนึ่งใน ปี 2006 และเริ่มต้นรอบที่สอง (second cycle) ในปี 2009 และจะครอบรอบในปี 2015 ทั้งนี้รูปแบบข้อสอบที่ใช้ในโครงการประเมินผลนักเรียนนานาชาติ (PISA) ปี 2009 ประกอบด้วยลักษณะข้อสอบแบบต่างๆ เช่น ข้อสอบแบบเลือกตอบหรือแบบเลือกตอบเชิงซ้อน (complex multiple-choice) การเติมคำ เขียนตอบสั้นๆ สำหรับข้อสอบที่มีการตรวจให้คะแนนทววิภาค เช่น ข้อสอบประเภทเลือกตอบและข้อสอบแบบตอบอิสระถ้าตอบข้อสอบได้ถูกต้องจะได้คะแนนเต็ม 1 คะแนน ถ้าตอบผิดจะได้ 0 คะแนน ส่วนข้อสอบที่มี

การตรวจให้คะแนนพหุวิภาค เช่น ข้อสอบเลือกตอบเชิงซ้น และข้อสอบแบบเปิด ถ้าตอบข้อสอบได้ถูกต้องจะได้คะแนนเต็ม 2 คะแนน ถ้าตอบถูกบางส่วนจะได้ 1 คะแนน และถ้าตอบผิดจะได้ 0 คะแนน นอกจากข้อสอบที่นักเรียนจะต้องผ่านการประเมินโครงการประเมินผลนักเรียนนานาชาติยังมีการเก็บแบบสอบถามซึ่งเป็นข้อมูลพื้นฐานทั้งของนักเรียนและสถานศึกษา ซึ่งจะเกิดขึ้นภายหลังจากนักเรียนทำแบบทดสอบเสร็จเรียบร้อย โดยการเก็บข้อมูลดังกล่าวยังสามารถนำมาเป็นข้อมูลที่ใช้สำหรับการกำหนดนโยบายและพัฒนาสถานศึกษาภายหลังการทดสอบอีกด้วย

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับโครงการประเมินผลนักเรียนนานาชาติมีการศึกษาหลากหลายด้านซึ่งส่วนใหญ่ที่มีการนำเสนอในงานวิจัยครั้งนี้มักเป็นการศึกษาที่เน้นที่การศึกษาการทำหน้าที่ต่างกันของข้อสอบ เช่น งานวิจัยของ Le (2009) Liu และ Wilson (2009) Yildirim และ Berberoglu (2009) Yildirim และ Yildirim (2011) Oliveri และคณะ (2012) และ Kankaras และ Moors (2014) ดังนั้น จากการทบทวนเอกสารและงานวิจัยดังกล่าวทำให้ผู้วิจัยสนใจที่จะศึกษาข้อมูลจากโครงการประเมินผลนักเรียนนานาชาติ (PISA) ปี 2009 ด้านการรู้เรื่องวิทยาศาสตร์ ในบริบทของประเทศไทย โดยการศึกษาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ การทำหน้าที่ต่างกันเป็นขั้น และการทำหน้าที่ต่างกันของแบบสอบ เพื่อการขยายแนวคิดและพัฒนาองค์ความรู้ทางด้านการทำหน้าที่ต่างกันของข้อสอบ การทำหน้าที่ต่างกันเป็นขั้น และการทำหน้าที่ต่างกันของแบบสอบสำหรับข้อสอบที่มีการตรวจให้คะแนนทั้งทวิวิภาคและพหุวิภาคหรือแบบผสมผสานภายในฉบับเดียวกัน (mixed format tests)

ตอนที่ 4 แนวคิดเกี่ยวกับมูลค่าเพิ่มของการจัดการศึกษาและการวิเคราะห์พหุระดับ

ในตอนที่สี่ผู้วิจัยแบ่งการนำเสนอออกเป็น 4 ส่วนได้แก่ มูลค่าเพิ่มของการจัดการศึกษา ขั้นตอนและวิธีการที่ใช้วิเคราะห์มูลค่าเพิ่ม การวิเคราะห์พหุระดับ และงานวิจัยที่เกี่ยวกับตัวแปรที่มีอิทธิพลต่อคุณภาพการจัดการศึกษาและการใช้โมเดลมูลค่าเพิ่มเพื่อประเมินคุณภาพการจัดการศึกษาโดยมีรายละเอียดดังต่อไปนี้

4.1 มูลค่าเพิ่มของการจัดการศึกษา

มูลค่าเพิ่ม (value-added) เป็นแนวคิดที่ได้มีการพัฒนามาจากเศรษฐศาสตร์ ถูกใช้เป็นเครื่องมือวัดผลการดำเนินงานทางธุรกิจในเชิงเศรษฐศาสตร์ มาจากคำว่า economic value added (EVA) ถูกพัฒนาขึ้นในช่วงปลายทศวรรษ 1980 โดยบริษัทที่ปรึกษาอเมริกัน stern sewart consulting group ทั้งนี้ EVA ถูกนำมาใช้เป็นตัววัดทางด้านการเงินและเป็นเกณฑ์ในการวัดผลการดำเนินงาน โดยคำนึงถึงผลต่างระหว่างมูลค่าของผลผลิต (output) กับมูลค่าของ

ปัจจัยนำเข้า (input) ภายหลังแนวคิดดังกล่าวได้ถูกนำมาใช้ทางการศึกษาเพื่ออธิบายการเพิ่มคุณค่า (value) ของสถานศึกษา ซึ่งมูลค่าเพิ่มสามารถทำนายจากพื้นฐานของนักเรียนและความสำเร็จเดิมของนักเรียน เช่น เพศ เชื้อชาติ เศรษฐานะ ผลสัมฤทธิ์ทางการเรียนเดิม ซึ่งไม่รวมผลที่เกิดจากปัจจัยที่อยู่นอกเหนือการควบคุมของสถานศึกษา โดยการทุ่มเทของสถานศึกษาทำให้เกิดความก้าวหน้าของนักเรียนซึ่งเป็นไปตามวัตถุประสงค์การศึกษาอันจะนำไปสู่ผลการเรียนรู้ของนักเรียน (Dawnes & Vindurampulle, 2007; Lambert & Line, 2000; Lissitz et al., 2006; OECD, 2008)

วิธีการวัดมูลค่าเพิ่มได้มีการนำมาใช้หลายประเทศเพื่อใช้ผลมาพิจารณาการดำเนินการของสถานศึกษาที่ส่งผลต่อผลสำเร็จของนักเรียน ดังตัวอย่างของระบบการประเมินในรัฐเทนเนสซีของประเทศสหรัฐอเมริกาที่มีชื่อระบบการประเมินว่าระบบการวัดมูลค่าเพิ่มของเทนเนสซี (tennessess value-added assessment system: TVAAS) ซึ่งออกแบบเพื่อวัดผลกระทบของสถานศึกษาและครูที่ส่งผลต่อความสำเร็จของนักเรียน ซึ่งพัฒนาโดย William L. Sanders และมีการศึกษาเพิ่มเติมไปสู่การดำเนินการระบบการประเมินมูลค่าเพิ่มทางการศึกษา (education value-added assessment system: EVAAS) นอกจากนี้ Heck ยังศึกษาการตรวจสอบผลกระทบของคุณภาพสถานศึกษาต่อผลลัพธ์ของสถานศึกษาและการพัฒนา โดยการประยุกต์ใช้วิธีการมูลค่าเพิ่ม โดยให้ความสำคัญไปที่ระดับรัฐเพื่อติดตามผลสำเร็จของนักเรียน และทำความเข้าใจเกี่ยวกับความสามารถรับผิดชอบของสถานศึกษา การรายงานผลการปฏิบัติงานของสถานศึกษาแก่ผู้กำหนดนโยบายบุคลากรของสถานศึกษาและผู้ปกครองเพื่อใช้ผลการเปรียบเทียบสถานศึกษาให้ความยุติธรรมยิ่งขึ้น (Heck, 2000; OECD, 2008; Sanders & Horn, 1998)

ผลการประเมินมูลค่าเพิ่มมีความสำคัญกับการนำมาใช้กำหนดนโยบายที่สำคัญ 3 ด้าน คือ การริเริ่มการพัฒนาสถานศึกษา (school improvement initiatives) ระบบความสามารถในการตรวจสอบได้ของสถานศึกษา (systems of school accountability) และการเลือกสถานศึกษา (school choice) ในส่วนของการริเริ่มการพัฒนาสถานศึกษาเป็นสิ่งสำคัญที่การพัฒนากิจกรรมหรือชุดกิจกรรมที่ต้องการพัฒนาเป็นอันดับแรก ผลจากการวิเคราะห์ทางสถิติจะให้ข้อสารสนเทศเกี่ยวกับความสัมพันธ์ระหว่างปัจจัยนำเข้าและตัวบ่งชี้ของความสามารถของสถานศึกษาที่สามารถบอกถึงยุทธวิธีที่ควรปฏิบัติหรือไม่ควรปฏิบัติ อันจะนำไปสู่การกำหนดนโยบายและวางแผนนโยบาย การแบ่งสรรทรัพยากร ในส่วนของความสามารถในการตรวจสอบได้ของสถานศึกษา ผลจากการวิเคราะห์มูลค่าเพิ่มสามารถตรวจสอบการปฏิบัติงานและผลลัพธ์จากการปฏิบัติได้ทั้งหมด เช่น การให้ข้อมูลแก่ผู้ที่เกี่ยวข้องและสาธารณะเกี่ยวกับเงินภาษีได้อย่างมีประสิทธิภาพ ในส่วนของการเลือกสถานศึกษา ข้อมูลที่ได้จากการวิเคราะห์นี้จะมีจุดหมายไปที่ผู้ปกครองและครอบครัวของนักเรียนเกี่ยวกับความสามารถที่แตกต่างของสถานศึกษาเพื่อการตัดสิน

เลือกสถานศึกษา ผู้ปกครองสามารถเลือกสถานศึกษาที่ดีที่สุดและเหมาะสมกับแหล่งทรัพยากรที่สามารถเกิดขึ้นในสถานศึกษา (Hoxby, 2003; McKewen, 1995; OECD, 2008)

ข้อมูลที่ได้จากการวิเคราะห์มูลค่าเพิ่มมีบทบาทสำคัญสำหรับนโยบายการจัดสรรแหล่งทรัพยากร และกลยุทธ์การสอนที่จำเป็นในแต่ละสถานศึกษา การวัดมูลค่าเพิ่มมีความสำคัญหลายประการสำคัญใช้ในการตัดสินแก่ผู้ที่เกี่ยวข้องหลายฝ่าย (Downes & Vindurampulle, 2007; Drury & Doran, 2003; OECD, 2008) ดังนี้

1) การประเมินนโยบายและโครงการ (policy and program evaluation) ข้อมูลจากการวัดมูลค่าเพิ่มสามารถวิเคราะห์ได้ถึงการปฏิบัติงานของสถานศึกษา ที่สามารถเปรียบเทียบผลกระทบของสถานศึกษาเชิงสัมพันธ์กับสถานศึกษาอื่นที่พิจารณาร่วมกับปัจจัยอื่น เช่น คะแนนดิบ การประเมินครู การประเมินตนเองของสถานศึกษา การทบทวนการตัดสินและแนวโน้มของสถานศึกษา ข้อมูลที่ได้สามารถช่วยในการตัดสินใจเชิงนโยบาย (informing policy-making) ที่มีประสิทธิภาพต่อการกำหนดนโยบายในระยะเริ่มต้น เป็นประโยชน์ในการระบุสถานศึกษาที่มีประสิทธิภาพต่อการกำหนดนโยบายในระยะเริ่มต้น เป็นประโยชน์ในการระบุสถานศึกษาที่มีค่าเฉลี่ยสูงกว่าและต่ำกว่าค่าเฉลี่ยของความสามารถได้อย่างชัดเจน สถานศึกษาที่มีความสามารถสูงสามารถให้ตัวอย่างในการปฏิบัติที่ดี ช่วยให้วางนโยบายให้มีความสอดคล้องกันระหว่างนโยบายแหล่งทรัพยากร และยุทธวิธีการสอนกับคุณลักษณะของนักเรียนและสถานศึกษา

2) เครื่องมือหนึ่งสำหรับการพัฒนาสถานศึกษา (a tool for school improvement) เนื่องจากการวัดมูลค่าเพิ่มจะนำคุณลักษณะของนักเรียนเข้าไปพิจารณาและสามารถระบุถึงวิธีการของสถานศึกษาที่สามารถทำให้ผู้เรียนมีความสามารถตามที่ความคาดหวัง โดยอิงความสำเร็จของนักเรียน การวิเคราะห์มูลค่าเพิ่มจึงเป็นสิ่งที่สะท้อนถึงสิ่งแวดล้อมและการปฏิบัติการของสถานศึกษาเพื่อแปลความถึงผลลัพธ์ของบริบทของสถานศึกษาและชั้นเรียน เพื่อเป็นแนวทางในการพัฒนาสถานศึกษา การวัดมูลค่าเพิ่มจะสามารถชี้ได้ถึงปัจจัยที่มีส่งผลให้นักเรียนแสดงสามารถได้เป็นอย่างดีหรือสามารถแสดงได้ต่ำกว่าที่คาดหวัง นั้นคือสามารถช่วยในอิทธิผลโดยตรงเพื่อพัฒนาผลลัพธ์การเรียนรู้ของนักเรียน

3) ประสิทธิผลของครู (teacher effectiveness) วิธีการวิเคราะห์มูลค่าเพิ่มมีผลต่อผลสัมฤทธิ์ของนักเรียนซึ่งนำไปใช้ในการประเมินประสิทธิภาพของครู จากการศึกษาการวิเคราะห์มูลค่าเพิ่ม พบว่า ประสิทธิภาพของครูเป็นตัวบ่งชี้ที่สำคัญของความสัมพันธ์ของสถานศึกษาต่อการเรียนรู้ของนักเรียน นอกจากนี้จากข้อสังเกตพบว่าการวิเคราะห์มูลค่าเพิ่มจะเป็นเครื่องมือในการลดช่องว่างของผลสัมฤทธิ์ของนักเรียนในข้อได้เปรียบและข้อเสียเปรียบของนักเรียนนอกจากนี้ข้อมูลที่ได้จะเป็นการให้ข้อมูลย้อนกลับแก่ครูเกี่ยวกับวิธีการสอน มีความชัดเจนในเป้าหมายของการเรียนรู้ที่ส่งผลต่อการเปลี่ยนแปลงวิธีการสอนและสามารถใช้เป็นข้อมูลสำหรับครูในการพัฒนา

นักเรียนต่อไป กล่าวคือสามารถระบุคุณลักษณะของนักเรียนที่ต้องมีการพัฒนาทักษะต่อไป และใช้เป็นข้อมูลต่อการวางแผนการพัฒนาของครู ตัวอย่างเช่น การวิเคราะห์มูลค่าเพิ่มจะระบุครูที่มีประสิทธิภาพในการยกระดับความสามารถของนักเรียนที่มีจุดอ่อน

4) เครื่องมือหนึ่งสำหรับความรับผิดชอบที่ตรวจสอบได้ (a tool for accountability) ตัวบ่งชี้การปฏิบัติงานของสถานศึกษาเป็นส่วนหนึ่งของการตรวจสอบได้ของสถานศึกษา การวัดมูลค่าเพิ่มจะเพิ่มความรอบครอบในการใช้ที่เป็นพื้นฐานสำหรับการกำหนดภาพรวมของ การปฏิบัติงานสถานศึกษา จากรายงานของ OECD (2008) เกี่ยวกับวิธีการมูลค่าเพิ่มจะได้รับข้อสรุปที่คล้ายกันเกี่ยวกับกระบวนการการศึกษา โดยรายงานว่าโมเดลมูลค่าเพิ่มเป็นวิธีที่ดีที่สุดที่แนะนำได้ว่าเป็นเครื่องมือที่ส่งเสริมการพัฒนาสถานศึกษา

5) การรายงานต่อผู้ปกครองและชุมชน (reporting to parents and community) การวัดมูลค่าเพิ่มจะให้ข้อสารสนเทศกับผู้ปกครองได้อย่างถูกต้องเกี่ยวกับวิธีการที่สถานศึกษาได้พัฒนาผลการเรียนรู้ของนักเรียน โดยรายงานให้ทราบถึงการปฏิบัติของสถานศึกษาในบริบทของสมรรถนะของนักเรียนและสิ่งแวดล้อมในชุมชนของพวกเรา การวัดมูลค่าเพิ่มจะให้ข้อสารสนเทศแก่ผู้ปกครองต่อผลสำเร็จของครูในการยกระดับผลสัมฤทธิ์ของนักเรียน

4.2 ขั้นตอนและวิธีการที่ใช้วิเคราะห์มูลค่าเพิ่ม

ขั้นตอนการวิเคราะห์มูลค่าเพิ่มแบ่งเป็น 5 ขั้นตอน คือ 1) การกำหนดเป้าหมายการวัดผลผลิตที่ต้องการเปรียบเทียบ 2) กำหนดชนิดของข้อมูลและการรวบรวมข้อมูล 3) การสร้างโมเดลการวิเคราะห์ซึ่งประกอบด้วยปัจจัยที่มีความสัมพันธ์อย่างมีเหตุผลกับผลผลิตที่เกิดขึ้น 4) การกำหนดหรือการสร้างเครื่องมือการวัดสถานะของนักเรียนตามปัจจัยที่กำหนดขึ้น 5) การเลือกใช้เทคนิคทางสถิติที่เหมาะสม นอกจากนี้ OECD ได้เสนอขั้นตอนที่สำคัญในการวิเคราะห์มูลค่าเพิ่มเพื่อสามารถนำไปปฏิบัติได้อย่างรวดเร็วและง่ายต่อการทำความเข้าใจ จำนวน 8 ขั้นตอน ซึ่งมีแนวทางเดียวกันกับขั้นตอนที่เสนอไปข้างต้น แต่มีการนำเสนอขั้นตอนที่ละเอียดเพิ่มมากขึ้นเกี่ยวกับแนวทางในการนำโมเดลมูลค่าเพิ่มไปดำเนินการใช้ซึ่งเสนอในขั้นที่ 5 6 7 และ 8 ดังนี้

ขั้นตอนที่ 1 การกำหนดวัตถุประสงค์และการวัดผลการปฏิบัติงานของสถานศึกษา ซึ่งขั้นตอนนี้จะกำหนดวัตถุประสงค์สำหรับการนำโมเดลมูลค่าเพิ่มไปใช้ ประกอบด้วยการกำหนดลักษณะของสารสนเทศของมูลค่าเพิ่มและอธิบายถึงการอธิบายคะแนนมูลค่าเพิ่ม การพิจารณาโครงสร้างของประเมินนักเรียนที่มีความสอดคล้องกับวัตถุประสงค์โมเดลมูลค่าเพิ่ม ตลอดจนการระบุเนื้อหาและระดับชั้นที่ประเมิน การกำหนดเป้าหมายการประเมิน และการพิจารณาศักยภาพของเครื่องมือ

ประเมินจะต้องมั่นใจว่าสามารถใช้ในโมเดลมูลค่าเพิ่ม ซึ่งเป็นสิ่งสำคัญที่จะนำมาใช้สำหรับนำมาแปลความหมายของผลการปฏิบัติ

ขั้นตอนที่ 2 การนำเสนอและการใช้ข้อสารสนเทศมูลค่าเพิ่ม เป็นการตัดสินและเลือกวิธีการที่เหมาะสมสำหรับการแสดงสารสนเทศของมูลค่าเพิ่ม ตลอดจนแนวทางการแปลความหมายของมูลค่าเพิ่ม ซึ่งสามารถสร้างขึ้นได้โดยผู้ที่เกี่ยวข้องในการจัดทำนโยบาย

ขั้นตอนที่ 3 การพิจารณาคุณภาพของข้อมูล เป็นขั้นตอนที่ควรจะพิจารณาระบบของข้อมูลในสถานศึกษาและสารสนเทศเพื่อรวบรวมสำหรับนำไปใช้ในโมเดลมูลค่าเพิ่ม ตลอดจนการกำหนดกลุ่มตัวอย่างนักเรียนที่เพียงพอต่อการนำมาใช้ในโมเดลมูลค่า ซึ่งอาจเกิดการสูญหายข้อมูล (missing data) ที่อาจเกิดการย้ายสถานศึกษาในช่วงที่มีการประเมิน ข้อมูลที่จำเป็นจะต้องสนับสนุนโมเดล หรือการนำไปใช้สำหรับการกำหนดนโยบายและพัฒนาโครงการต่อไป

ขั้นตอนที่ 4 การเลือกโมเดลมูลค่าเพิ่มที่เหมาะสม การกำหนดวัตถุประสงค์เชิงนโยบายและการใช้คะแนนมูลค่าเพิ่มของสถานศึกษาจะเป็นแนวทางในการเลือกโมเดล สิ่งจำเป็นที่จะต้องกำหนดคือการเลือกใช้สถิติและวิธีการวิเคราะห์ภายใต้โมเดลมูลค่าเพิ่ม ความแปรปรวนในโมเดลมูลค่าเพิ่ม เพื่อประเมินความเหมาะสมของลักษณะของโมเดล ลักษณะเฉพาะของโมเดลอาจจะทำให้เกิดความแตกต่างของสถานศึกษา ผลกระทบของการเลือกโมเดลขึ้นอยู่กับความแตกต่างของสถานศึกษาในจุดประสงค์การใช้คะแนนมูลค่าเพิ่ม

ขั้นตอนที่ 5 กลยุทธ์กระตุ้นการสื่อสารและผู้ที่มีส่วนได้ส่วนเสีย เป้าหมายของวิธีการสื่อสารจะสอดคล้องกับวัตถุประสงค์ของนโยบาย การวัดความสามารถการปฏิบัติงานของโรงเรียนจะนำมาตัดสินควรจะมีการอธิบายได้อย่างชัดเจน การกระตุ้นและการสื่อสารต่อผู้ที่เกี่ยวข้องจะต้องชี้แจงได้อย่างชัดเจน เพื่อนำมาใช้แปลความหมายในผลของมูลค่าเพิ่งของสถานศึกษา

ขั้นตอนที่ 6 การฝึกอบรม เป้าหมายของการฝึกอบรมมีเป้าหมายในการพัฒนากลุ่มผู้ใช้การฝึกอบรมผู้บริหารสถานศึกษาและครู จะเป็นการอธิบายเกี่ยวกับผลของมูลค่าเพิ่มและสามารถนำมาใช้ภายในสถานศึกษาเพื่อวัตถุประสงค์ของการพัฒนาสถานศึกษา รวมถึงการอบรมเกี่ยวกับการวิเคราะห์ทางสถิติและการใช้ข้อสารสนเทศที่จำเป็น นอกจากนี้การฝึกอบรมสำหรับผู้ปกครองมีเป้าหมายสำหรับเพื่อการแปลความหมายของคะแนนมูลค่าเพิ่มเพื่อความสะดวกในการเลือกสถานศึกษา

ขั้นตอนที่ 7 การนำร่องโครงการ เป็นการตัดสินใจเกี่ยวกับขนาดและคุณลักษณะของกลุ่มตัวอย่างของสถานศึกษาที่จะมีความสอดคล้องกับวัตถุประสงค์ของนโยบายการนำระบบของโมเดลมูลค่าเพิ่มไปใช้ การนำร่องโครงการควรจะใช้เพื่อประเมินกิจกรรมที่มีความเชื่อมโยงกับโมเดลมูลค่าเพิ่ม

ขั้นตอนที่ 8 การนำไปพัฒนา เป็นขั้นตอนสุดท้ายของการนำโมเดลมูลค่าเพิ่มไปใช้ ซึ่งเป็นการติดตามผลของคุณภาพ แหล่งทรัพยากร และมีการพิจารณาถึงความสอดคล้องของโมเดลมูลค่าเพิ่มกับนโยบาย

4.3 การวิเคราะห์พหุระดับ

ข้อมูลการวิจัยทางสังคมศาสตร์ส่วนใหญ่มีลักษณะเป็นพหุระดับ เนื่องจากธรรมชาติข้อมูลมีความเกี่ยวข้องอยู่หลายระดับ โดยมีความสัมพันธ์ในลักษณะต่างๆ นอกจากนี้การที่จะได้ผลการวิเคราะห์ที่ถูกต้องจำเป็นต้องคำนึงระดับข้อมูลที่จะนำมาใช้ในการวิเคราะห์ (ศิริชัย กาญจนวาสี, 2550ข) ดังนี้

1) การจัดองค์กรและข้อมูลขององค์กร การจัดองค์กรโดยทั่วไปและการแบ่งส่วน การบริหารงานภายในองค์กรมีลักษณะการจัดตั้งเป็นลำดับขั้น (hierarchy) หน่วยงานหรือการบริหารระดับสูงต้องมีความรับผิดชอบหรือมีอิทธิพลต่อหน่วยงาน หรือการบริหารในระดับรองลงมาเป็นลำดับขั้น การจัดองค์กรและการดำเนินงานทางการศึกษาก็เช่นกัน ตัวแปรในระดับบนย่อมส่งผลกระทบต่อตัวแปรในระดับล่างเป็นทอดๆ ไป เช่น นโยบายของกระทรวงศึกษาธิการย่อมส่งผลต่อการบริหารการศึกษาระดับเขตการศึกษา การบริหารการศึกษาระดับโรงเรียน การเรียน การสอนในชั้นเรียน ผลสัมฤทธิ์ทางการเรียนของนักเรียน เป็นต้น

2) ปฏิสัมพันธ์ระหว่างตัวแปรภายในองค์กร เนื่องจากการจัดองค์กรและธรรมชาติของข้อมูลในองค์กรมีลักษณะเป็นพหุระดับมีความสลับซับซ้อนและไม่หยุดนิ่ง (dynamic) ดังนั้นตัวแปรต่างๆ ที่อยู่ในระดับเดียวกันและต่างระดับจึงมีปฏิสัมพันธ์กัน (interaction) ซึ่งกันและกันตลอดเวลา ซึ่งจะต้องมีการศึกษาทำความเข้าใจความสัมพันธ์ระหว่างตัวแปรในระดับเดียวกันและปฏิสัมพันธ์ระหว่างตัวแปรต่างระดับ

3) การรวมกลุ่มสมาชิกภายในองค์กร มักเป็นหน่วย แผนก ฝ่าย โดยมีจุดมุ่งหมายใดจุดมุ่งหมายหนึ่ง หน่วยทางการศึกษาอาจเป็นชั้นเรียน โรงเรียน กลุ่มโรงเรียน โรงเรียนในเขตพื้นที่การศึกษา เป็นต้น ดังนั้นหน่วยงานทางการศึกษาจึงประกอบด้วยสมาชิก (ผู้เรียน ผู้สอน ผู้บริหาร ฯลฯ) มาอยู่รวมกันด้วยจุดมุ่งหมายเฉพาะอย่างไม่ได้มาอยู่รวมกันอย่างสุ่ม (random) ดังนั้นการศึกษาเปรียบเทียบระหว่างหน่วยงานทางการศึกษาจะต้องมั่นใจได้ว่ามีการควบคุมความแตกต่างเบื้องต้นของหน่วย (pre-existing difference) และคัดเลือกตัวแปรในระดับต่างๆ ที่เกี่ยวข้องได้อย่างครอบคลุม

1) การวิเคราะห์พหุระดับด้วยโมเดลเชิงเส้นตรงระดับลดหลั่น

การสร้างโมเดลสมการทำนายตัวแปรตามที่มีประสิทธิภาพ โดยเฉพาะอย่างยิ่งการวิเคราะห์สมการทำนายแบบพหุระดับจะนำไปสู่ข้อค้นพบหรือสารสนเทศที่มีความถูกต้อง (multi level analysis) เป็นเทคนิคทางสถิติที่ใช้ในการวิเคราะห์อิทธิพลของตัวแปรทำนายหลายระดับที่มีต่อตัวแปรตาม ซึ่งเทคนิคทำนายมีโครงสร้างเป็นระดับลดหลั่นอย่างน้อย 2 ระดับ โดยตัวแปรทำนายและตัวแปรตามที่อยู่ระดับล่างต่างมีความสัมพันธ์ซึ่งกันและกันและได้รับอิทธิพลร่วมกันจากตัวแปรทำนายที่อยู่บน (ศิริชัย กาญจนวาสี, 2550ข) ในการวิเคราะห์ข้อมูลพหุระดับนั้นการวิเคราะห์ประมาณค่าส่วนประกอบความแปรปรวน (analysis of variance component estimation) โดยที่ตัวแปรที่วัดได้ในระดับนักเรียนมีความแปรปรวนซึ่งแยกส่วนประกอบได้ตามระดับที่ลดหลั่นกัน เช่น กรณีมีสามระดับ คือระดับนักเรียน ระดับห้องเรียนและระดับโรงเรียน และแสดงส่วนประกอบความแปรปรวนได้ดังนี้

$$\sigma_y^2 = \sigma_{pupil}^2 + \sigma_{class}^2 + \sigma_{school}^2$$

โดยที่

σ_y^2 แทน ความแปรปรวนของตัวแปรตามที่สนใจศึกษา

σ_{pupil}^2 แทน ความแปรปรวนระหว่างนักเรียนภายในโรงเรียน

σ_{class}^2 แทน ความแปรปรวนระหว่างห้องเรียนภายในโรงเรียน

σ_{school}^2 แทน ความแปรปรวนระหว่างโรงเรียน

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2) การประมาณค่าพารามิเตอร์ในโมเดลพหุระดับ

การประมาณค่าพารามิเตอร์ในโมเดลพหุระดับสามารถกระทำได้หลายวิธี โดยวิธีที่สำคัญ 2 วิธี ได้แก่ วิธีกำลังสองน้อยที่สุด (OLS) และวิธีตามโมเดล HLM (hierarchical linear model approach) (Raudenbush & Bryk, 1986 อ้างถึงใน ศิริชัย กาญจนวาสี, 2550ข)

1. OLS (ordinary least square approach) เป็นการใช้เทคนิคกำลังสองน้อยที่สุดแบบธรรมดา ซึ่งนิยมใช้กันทั่วไปในโปรแกรมคอมพิวเตอร์มาตรฐานแบบประเพณีนิยมทั้งหลายที่มีอยู่ในปัจจุบัน แต่มีข้อเสียในด้านความเหมาะสมของโมเดลที่ใช้วิเคราะห์และผลการวิเคราะห์ที่ได้รับ ตลอดจนมีความยุ่งยากในการเตรียมข้อมูลสำหรับการวิเคราะห์

2. โมเดล HLM (hierarchical linear model approach) เป็นการใช้เทคนิคการวิเคราะห์ความแปรปรวนแบบผสม ใช้หลักการสัมประสิทธิ์แบบสุ่มและการประมาณ

ค่าพารามิเตอร์ด้วยวิธีของเบย์ส (bayesian estimation) เทคนิค HLM พัฒนาโดย Raudenbush & Bryk ที่จะให้ผลการวิเคราะห์พหุระดับที่มีความคงเส้นคงวาและน่าเชื่อถือมากกว่าวิธี OLS

3) จุดเด่นของเทคนิควิธีของ HLM

เทคนิควิธีของ HLM มีจุดเด่นที่สำคัญ ดังนี้

1) สามารถตรวจสอบความเหมาะสมของโมเดล (accuracy of a model) เพื่อพิจารณาถึงโมเดล 2 ระดับ ได้แก่

- Within-Group Model

  $Y_{ij} = B_{Oj} + B_{1j}(X_{ij}) + R_{ij}$

- Between-Group Model

  $B_{Oj} = G_{00} + G_{01}(Z_j) + U_{0j}$

  $B_{1j} = G_{10} + G_{11}(Z_j) + U_{1j}$

โมเดลดังกล่าวสามารถมีตัวทำนายระดับบุคคล (X_{ij}) และตัวแปรทำนายระดับกลุ่ม (Z_j) ได้หลายตัว และการทำนายตัวแปรตามระหว่างหน่วยไม่จำเป็นต้องใช้ Z_j ชุดเดียวกันก็ได้จากโมเดล 2 ระดับดังกล่าว HLM สามารถตรวจสอบความเหมาะสมของโมเดลได้ดังนี้

1.1) ตัวแปร X_{ij} ส่งผลอย่างมีนัยสำคัญทางสถิติต่อ Y_{ij} หรือไม่

HLM จะคำนวณผลเฉลี่ยของ X_{ij} ที่มีต่อ Y_{ij} จากทุกกลุ่มและทดสอบนัยสำคัญทางสถิติ

1.2) ค่าพารามิเตอร์ของแต่ละหน่วยมีความผันแปรระหว่างหน่วยหรือไม่

HLM จะแบ่งผลของพารามิเตอร์ของแต่ละกลุ่มออกเป็นอิทธิพลคงที่ (fixed effect) และอิทธิพลสุ่ม (random effect) ดังนี้

$$B_{Oj} = G_{00} + U_{0j}$$

$$B_{1j} = G_{10} + U_{1j}$$

(Fixed effect) (Random effect)

(ค่าเฉลี่ย) (ค่าความคลาดเคลื่อน)

HLM จะทำการทดสอบค่าเฉลี่ย fixed effect ว่าเป็น 0 หรือไม่โดยใช้สถิติการทดสอบที (t-test)

$$H_0: G_{00} = 0$$

$$H_1: G_{00} \neq 0$$

และ

H_0: $G_{10} = 0$

H_1: $G_{10} \neq 0$

ถ้า fixed effect ไม่เป็น 0 แสดงว่า ค่าคงที่ (intercept) มีผลต่อ Y_{ij} หรือ $B1_j$ แต่ถ้ามีค่าเป็น 0 แสดง ค่าคงที่ (intercept) ดังกล่าวไม่มีผลต่อ Y_{ij} หรือ $B1_j$ นอกจากนี้ HLM ยังทำการทดสอบความแปรปรวนของ random effect หรือ parameter variance ว่าเป็น 0 หรือไม่ โดยใช้สถิติทดสอบไคสแควร์ $(\chi^2 - \text{test})$

H_0: $Var(B_{0j})$ or $Var(U_{0j}) = 0$

H_1: $Var(B_{0j})$ or $Var(U_{0j}) > 0$

และ

H_0: $Var(B_{1j})$ or $Var(U_{1j}) = 0$

H_1: $Var(B_{1j})$ or $Var(U_{1j}) > 0$

ถ้า random effect ไม่เป็น 0 แสดงว่า พารามิเตอร์ (B_{Oj}) และ (B_{1j}) มีความผันแปรระหว่างกลุ่ม แต่ถ้ามีค่าเป็น 0 แสดงว่า พารามิเตอร์ดังกล่าวไม่มีความผันแปรระหว่างกลุ่ม ซึ่งสามารถตั้งข้อจำกัดให้เป็นค่าคงที่ ในการวิเคราะห์ได้ในกรณีที่ค่าพารามิเตอร์ของแต่ละกลุ่มมีความผันแปรระหว่างกลุ่ม จึงสมเหตุสมผลที่จะหาตัวแปรทำนายระหว่างกลุ่มมาอธิบายความผันแปรดังกล่าว

Chulalongkorn University

2) สามารถประมาณค่าพารามิเตอร์ได้อย่างคงเส้นคงวาและน่าเชื่อถือ

จากสมการระดับที่ 1 (within – group Model) และระดับที่ 2 (between – group model) การแจกแจงของค่าความคลาดเคลื่อนหรือค่าส่วนที่เหลือเป็นดังนี้

$R_{ij} \sim N(0, \tau^2)$

$U0_j \sim N(0, \tau_0)$

$U1_j \sim N(0, \tau_1)$

4) การประยุกต์ใช้โมเดลเชิงเส้นแบบลดหลั่นสำหรับการวิเคราะห์มูลค่าเพิ่ม

1) การตรวจสอบความเหมาะสมของโมเดล

ความเหมาะสมของโมเดลพิจารณาในโมเดล 2 ระดับ (ศิริชัย กาญจนวาสี, 2550ข) ดังนี้

โมเดลภายในกลุ่ม (within-group model)

$$Y_{ij} = B_{0j} + B_{1j}X_{ij} + R_{ij}$$

โมเดลระหว่างกลุ่ม (between-group model)

$$B_{0j} = G_{00} + G_{01}Z_j + U_{0j}$$

$$B_{1j} = G_{10} + G_{11}Z_j + U_{1j}$$

การตรวจสอบความเหมาะสมของโมเดลทำได้ดังนี้

1) ตัวแปร X_{ij} ส่งผลอย่างมีนัยสำคัญทางสถิติต่อ Y_{ij} หรือไม่ HLM จะคำนวณผลเฉลี่ยของ X_{ij} ที่มีต่อ Y_{ij} จากทุกกลุ่มและทดสอบนัยสำคัญทางสถิติ

2) ค่าพารามิเตอร์ของแต่ละหน่วยมีความผันแปรระหว่างหน่วยหรือไม่ HLM จะแบ่งผลของพารามิเตอร์ของแต่ละกลุ่มออกเป็นอิทธิพลคงที่ (fixed effect) คือ G_{00}, G_{01} และอิทธิพลสุ่ม (random effect) คือ U_0j, U_{1j} ดังนี้

$$B_{0j} = G_{00} + U_{0j}$$

$$B_{1j} = G_{10} + U_{1j}$$

การทดสอบค่าเฉลี่ยของอิทธิพลคงที่ว่าเป็นศูนย์หรือไม่ โดยใช้สถิติทดสอบที (t-test) มีสมมติฐานดังนี้

$$H_0: G_{00} = 0$$

$$H_1: G_{00} \neq 0$$

และ

$$H_0: G_{10} = 0$$

$$H_1: G_{10} \neq 0$$

ถ้าอิทธิพลคงที่ (fixed effect) ไม่เป็นศูนย์แสดงว่าคงที่หรือค่าตัดแกน (intercept) มีผลต่อ Y_{ij} หรือ B_{1j} แต่ถ้ามีค่าเป็นศูนย์ แสดงว่าค่าคงที่ไม่มีผลต่อ Y_{ij} หรือ B_{1j}

การทดสอบความแปรปรวนของอิทธิพลสุ่มว่าเป็นศูนย์หรือไม่ โดยใช้สถิติทดสอบไคสแควร์ (χ^2-test) มีสมมติฐานดังนี้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

H_0: $Var(B_{0j})$, $Var(U_{0j}) = 0$, H_1: $Var(B_{0j})$, $Var(U_{0j}) > 0$

และ

H_0: $Var(B_{1j})$, $Var(U_{1j}) = 0$, H_1: $Var(B_{1j})$, $Var(U_{1j}) > 0$

ถ้าอิทธิพลสุ่ม (random effect) ไม่เป็นศูนย์แสดงว่าค่า $B_{O}j$ และ B_{1j} มีความผันแปรระหว่างกลุ่ม แต่ถ้าไม่เป็นศูนย์แสดงว่า ค่า B_{0j} และ B_{1j} ไม่มีความผันแปรระหว่างกลุ่ม และพบว่าค่าพารามิเตอร์ B_{0j} และ B_{1j} มีความผันแปรระหว่างกลุ่มจึงสามารถนำตัวแปรทำนายระหว่างกลุ่มมาทำการอธิบายความผันแปรดังกล่าว

2) ขั้นตอนการวิเคราะห์

การวิเคราะห์ใน HLM จะเริ่มจากขั้นตอนที่ 1 ที่ยังไม่มีเงื่อนไขหรือตัวแปรทำนายในระดับที่ 1 และเพื่อเป็นการทดสอบค่าเฉลี่ยของตัวแปรตามว่ามีค่าแตกต่างจากศูนย์หรือไม่ ขั้นตอนที่ 2 เป็นการนำตัวแปรทำนายระดับบุคคลเข้ามาวิเคราะห์ และขั้นตอนที่ 3 จะนำตัวแปรทำนายระดับบุคคลและระดับกลุ่มเข้ามาวิเคราะห์ร่วมด้วย ดังนี้

ขั้นตอนที่ 1 วิเคราะห์ขั้นโมเดลศูนย์ (null model) หรือแบบไม่มีเงื่อนไขอย่างสมบูรณ์ (fully uncoditional models) เป็นการวิเคราะห์ขั้นแรกเพื่อให้เห็นภาพรวมของตัวแปรเกณฑ์หรือตัวแปรตาม โดยไม่นำตัวแปรทำนายเข้ามาร่วมพิจารณา และเพื่อตรวจสอบว่าตัวแปรเกณฑ์มีความแปรปรวนในหน่วยและระหว่างหน่วยเพียงพอที่จะวิเคราะห์ตัวแปรทำนายที่มีอิทธิพลในขั้นตอนต่อไปหรือไม่ มีสมการวิเคราะห์ดังนี้

ระดับที่ 1 ภายในกลุ่ม

$$Y_{ij} = B_{0j} + R_{ij}$$

ระดับที่ 2 ระหว่างกลุ่ม

$$B_{0j} = G_{00} + U_{0j}$$

ขั้นตอนที่ 2 วิเคราะห์ขั้นโมเดลอย่างง่าย (simple model) หรือแบบไม่มีเงื่อนไข (unconditional models) เป็นการวิเคราะห์โดยการนำตัวแปรทำนายระดับนักเรียนเข้ามาวิเคราะห์ เพื่อดูว่าตัวแปรทำนายเหล่านั้นมีอิทธิพลหรือไม่ มีสมการวิเคราะห์ดังนี้

ระดับที่ 1 ภายในกลุ่ม

$$Y_{ij} = B_{0j} + B_{1j}X_{ij} + R_{ij}$$

ระดับที่ 2 ระหว่างกลุ่ม

$B_{0j} = G_{00} + U_{0j}$

$B_{1j} = G_{10} + U_{1j}$

ขั้นตอนที่ 3 วิเคราะห์ขั้นโมเดลสมมติฐาน (hypothetical model) โดยการนำตัวแปรทำนายระดับนักเรียน และพิจารณาว่ามีความเหมาะสมจากการวิเคราะห์ระดับนักเรียนมาวิเคราะห์ร่วมกับตัวแปรทำนายในระดับสถานศึกษา เพื่อตรวจสอบอิทธิพลของตัวแปรระดับสถานศึกษาที่มีต่อตัวแปรเกณฑ์ มีสมการวิเคราะห์ดังนี้

ระดับที่ 1 ภายในกลุ่ม

$Y_{ij} = B_{0j} + B_{1j}X_{ij} + R_{ij}$

ระดับที่ 2 ระหว่างกลุ่ม

$B_{0j} = G_{00} + G_{01}Z_j + U_{0j}$

$B_{1j} = G_{10} + G_{11}Z_j + U_{1j}$

3) สัมประสิทธิ์การทำนาย

สัมประสิทธิ์การทำนายของโมเดลแต่ละระดับ (R^2) หรือสัดส่วนความแปรปรวนที่อธิบายได้ (proportion variance explained) สามารถคำนวณได้จากการเทียบสัดส่วนความแปรปรวนค่าส่วนที่เหลือที่ลดลง เมื่อมีตัวแปรทำนายกับเมื่อไม่มีตัวแปรทำนาย มีสูตรดังนี้

R^2 = ความแปรปรวนของค่า residual ที่ลดลงเมื่อมีตัวแปรทำนาย / ความแปรปรวนของค่า residual เมื่อไม่มีตัวแปรทนาย

หรือ $$R^2 = \frac{\tau_{00}(fullyunconditional) - \tau_{00}(hypothetical)}{\tau_{00}(fullyuncontional)}$$

เมื่อ $\tau_{00}(fullyunconditional)$ และ $\tau_{00}(hypothetical)$ เป็นความแปรปรวนของค่าเศษเหลือในสมการที่ยังไม่ใส่ตัวแปรทำนาย และสมการที่ใส่ตัวแปรทำนายตามลำดับ ดังนั้น ความแปรปรวนทั้งหมดของตัวแปรตามที่อธิบายได้โดยโมเดล 2 ระดับ หาได้จากผลรวมของความแปรปรวนที่อธิบายได้ในระดับที่ 1 และระดับที่ 2

4.4 งานวิจัยที่เกี่ยวข้องกับปัจจัยหรือตัวแปรที่ส่งผลต่อคุณภาพการจัดการศึกษาและการใช้โมเดลมูลค่าเพิ่มเพื่อประเมินคุณภาพการจัดการศึกษา

จากการศึกษางานวิจัยที่เกี่ยวข้องกับปัจจัยหรือตัวแปรที่ส่งผลต่อคุณภาพการจัดการศึกษาและการใช้โมเดลมูลค่าเพิ่มเพื่อประเมินคุณภาพการจัดการศึกษา ผู้วิจัยแบ่งการนำเสนองานวิจัยที่เกี่ยวข้องออกเป็น 2 ประเด็น ประเด็นแรก คือ ปัจจัยหรือตัวแปรที่ส่งผลต่อคุณภาพการจัดการศึกษา และประเด็นที่สอง คือ การใช้โมเดลมูลค่าเพิ่มเพื่อประเมินคุณภาพการจัดการศึกษา รายละเอียดดังนี้

4.4.1 ปัจจัยหรือตัวแปรที่ส่งผลต่อคุณภาพการจัดการศึกษา

งานวิจัยและเอกสารที่เกี่ยวข้องกับปัจจัยหรือตัวแปรที่สงผลต่อคุณภาพการจัดการศึกษา โดยคำนึงถึงผลสัมฤทธิ์ทางการเรียนของนักเรียนมีอยู่หลายปัจจัยและหลายตัวแปรด้วยกัน ทั้งนี้ขึ้นอยู่กับบริบทของสถานศึกษาว่ามีความแตกต่างกันมากน้อยเพียงใด ผลกระทบต่อคุณภาพการจัดการศึกษาสามารถจำแนกออกได้เป็น 2 ส่วนหลัก ได้แก่ ระดับนักเรียนซึ่งเป็นบริบทที่เกี่ยวข้องโดยตรงและเป็นผลผลิตเป้าหมายที่ต้องการให้มีคุณภาพ โดยมีปัจจัยที่เกี่ยวข้องกับนักเรียน เช่น ทัศนคติของนักเรียน พื้นฐานของครอบครัว พื้นฐานความรู้เดิมของแต่ละคน เป็นต้น ส่วนระดับสถานศึกษาซึ่งเป็นสถาบันรองรับนักเรียนเกี่ยวข้องกับการบริหารจัดการที่ส่งผลต่อคุณภาพของนักเรียน โดยการสนับสนุนในเรื่องสิ่งอำนวยความสะดวกต่อการเรียนรู้ การพัฒนาครูหรือบุคคลากรทางการศึกษา แม้กระทั้งการกำหนดนโยบายหรือการวางแผนพัฒนานักเรียน ดังนั้นในการวิจัยครั้งนี้จึงแบ่งตัวแปรทางการศึกษาออกเป็น 2 ระดับ โดยคำนึงถึงข้อมูลพื้นฐานที่มีการศึกษาตามโครงการประเมินผลนักเรียนนานาชาติ ดังนี้

4.4.1.1 ตัวแปรระดับนักเรียน

การศึกษาเกี่ยวกับตัวแปรระดับนักเรียนที่ส่งผลต่อการเรียนรู้ของนักเรียน จากการศึกษาและทบทวนทั้งในประเทศและต่างประเทศ พบว่า งานวิจัยของ Caro (2009) ซึ่งศึกษาเศรษฐานะของครอบครัวและโอกาสความไม่เท่าเทียมกัน ชี้ให้เห็นว่า นักเรียนที่มาจากครอบครัวที่มีเศรษฐานะดีมักมีพัฒนาการทักษะการเรียนรู้รวดเร็วมากกว่านักเรียนที่มาจากครอบครัวที่มีเศรษฐานะไม่ดี ซึ่งผลที่ตามมาก็คือทำให้เกิดช่องว่างทางการเรียนรู้ เช่นเดียวกับ Sun และคณะ (2012) ซึ่งได้ศึกษาแนวคิดโมเดลพหุระดับเพื่อสำรวจปัจจัยที่มีผลกระทบต่อผลสัมฤทธิ์วิชาวิทยาศาสตร์ในโรงเรียนมัธยมศึกษาจากงานวิจัย พบว่า นักเรียนที่มีเศรษฐานะทางครอบครัวที่ดี มีแรงกระตุ้นสูง และมีความเชื่อในประสิทธิภาพของตน (Self-efficacy) มักจะมีแนวโน้มจะมี

ผลสัมฤทธิ์ทางวิทยาศาสตร์จากผลการทดสอบ PISA สูงตามไปด้วย Johnson และคณะ (2007) พบว่า เศรษฐานะมีส่วนเกี่ยวข้องกับผลสัมฤทธิ์ทางการเรียนของนักเรียน โดยที่เศรษฐานะของผู้ปกครองมีอิทธิผลต่อศักยภาพการบริหารของสถานศึกษา ส่วน Lee และ Wong (2004) ศึกษาผลกระทบการตรวจสอบในด้านความเท่าเทียมกันทางเชื้อชาติและเศรษฐานะโดยการพิจารณาทั้งแหล่งทรัพยากรของโรงเรียนและผลสัมฤทธิ์ทางการเรียน พบว่า เชื้อชาติและเศรษฐานะส่งผลต่อผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ตามนโยบายการตรวจสอบได้ของรัฐ (accountability policy) โดยอาศัยแนวคิดแบบการขับเคลื่อนตามสมรรถนะ (performance-driven approach) และ Rumberger และ Palardy (2005) ศึกษาผลกระทบองค์ประกอบที่เกี่ยวข้องกับผลสัมฤทธิ์ของนักเรียนโรงเรียนมัธยมศึกษา พบว่า ระดับเศรษฐานะเฉลี่ยของนักเรียนมีผลกระทบต่อพัฒนาการจากผลสัมฤทธิ์ทางการเรียน ส่วน Areepattamannil และ Kaur (2013) ในงานวิจัย พบว่า ภาษาที่ใช้สื่อสารในครอบครัว ความมั่งคั่งของครอบครัว และเศรษฐานะทางครอบครัวของนักเรียนที่มีถิ่นอาศัยภายในประเทศล้วนเป็นองค์ประกอบที่มีผลต่อผลสัมฤทธิ์ทางวิทยาศาสตร์ นอกจากนี้ยังมีนักวิชาการในประเทศไทยที่ดำเนินการศึกษาตัวแปรหรือปัจจัยที่ส่งต่อการเรียนรู้ เช่น รัชนก บุญปู่ (2547) ได้ศึกษาปัจจัยที่มีอิทธิผลต่อผลการเรียนรู้ตามมาตรฐานหลักสูตรการศึกษาขั้นพื้นฐานในโรงเรียนนำร่องหลักสูตรสถานศึกษาโดยการวิเคราะห์พหุระดับ จากผลการวิจัยชี้ให้เห็นว่า เจตคติต่อการเรียน พฤติกรรมการเรียน และความคาดหวังในการศึกษาต่อมีอิทธิพลต่อผลการเรียนรู้ตามมาตรฐานหลักสูตรการศึกษาขั้นพื้นฐานในโรงเรียนนำร่องหลักสูตรสถานศึกษา ส่วนทวีศิลป์ สารแสน (2543) ได้กล่าวไว้ว่า สภาพแวดล้อมทางการเรียนในห้องเรียนเป็นองค์ประกอบที่ทำให้นักเรียนเกิดความพอใจในการเรียน มีความกระตือรือร้นอยากเรียนรู้ ซึ่งการสร้างบรรยากาศให้เอื้อประโยชน์ต่อการเรียนรู้ของนักเรียนมักทำให้การเรียนการสอนในห้องเรียนมีประสิทธิภาพตามไปด้วย นอกจากนี้ศักดิ์ชัย จันทะแสง (2550) ยังได้เสนอแนะเพิ่มเติมว่าความถนัดทางด้านการเรียนด้านจำนวนส่งผลต่อผลสัมฤทธิ์ทางการเรียนวิทยาศาสตร์ เช่นเดียวกับ จารุวรรณ นาคคูบัว (2552) ชี้ให้เห็นว่าเชาว์ปัญญาด้านจำนวนส่งผลบวกต่อผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์

4.4.1.2 ตัวแปรระดับสถานศึกษา

การศึกษาเกี่ยวกับตัวแปรระดับสถานศึกษาที่ส่งผลต่อการเรียนรู้ของนักเรียน จากการศึกษาและทบทวนงานวิจัยทั้งในประเทศและต่างประเทศ พบว่า Lee และ Wong (2004) ศึกษาผลกระทบการตรวจสอบในด้านความเท่าเทียมกันทางเชื้อชาติและเศรษฐานะโดยการพิจารณาทั้งแหล่งทรัพยากรของโรงเรียนและผลสัมฤทธิ์ทางการเรียน ชี้ให้เห็นว่า เชื้อชาติและเศรษฐานะมีผลต่อร้อยละงบประมาณของรัฐที่จัดสรรให้สถานศึกษาแต่มีการจัดสรรเพิ่มขึ้นเพียงเล็กน้อยในเขตที่พบว่ามีโรงเรียนที่ยากจน ส่วนงานวิจัยของ Ryan (2012) ซึ่งศึกษาการจัดสรรทรัพยากรทางการ

ศึกษาในเขตพื้นที่การศึกษาของรัฐโรดไอแลนด์โดยการศึกษาความสัมพันธ์ระหว่างค่าใช้จ่ายและผลสัมฤทธิ์ทางการเรียนของนักเรียน พบว่า ความร่ำรวยของโรงเรียนมีความสัมพันธ์อย่างมีนัยสำคัญทางลบกับความสามารถของนักเรียน Lubienski และ Lubienski (2006) ศึกษาผลสัมฤทธิ์ทางการเรียนของโรงเรียนทางเลือก โรงเรียนเอกชน และโรงเรียนรัฐบาล พบว่า นักเรียนที่เรียนโรงเรียนเอกชนมีผลสัมฤทธิ์ทางการเรียนที่สูงกว่านักเรียนโรงเรียนของรัฐ และนักเรียนที่ด้อยโอกาสส่วนใหญ่มักจะเลือกเข้าศึกษาต่อในโรงเรียนของรัฐบาลมากกว่า และ Subedi (2003) ศึกษาปัจจัยที่มีอิทธิพลกับผลสัมฤทธิ์ทางการเรียนของนักเรียนระดับมัธยมศึกษาในประเทศเนปาล ชี้ให้เห็นว่าการใช้ทรัพยากรทางการศึกษามีต่อผลผลสัมฤทธิ์ทางการเรียนของนักเรียนภายในชั้นโดยขนาดชั้นเรียนมีอิทธิผลทางลบต่อระดับคะแนนผลสัมฤทธิ์ทางการเรียน นอกจากนี้ยังมีวิชาการในประเทศไทย ได้แก่ รัชนก บุญปู่ (2547) ชี้ให้เห็นว่า คุณภาพการสอนและระดับการศึกษาของผู้บริหารมีอิทธิพลต่อการจัดการศึกษาในระดับโรงเรียน นอกจากนี้ ทิศนา แขมมณี และคณะ (2545) ได้กล่าวถึงปัจจัยที่เอื้อต่อการพัฒนากระบวนการเรียนรู้ ได้แก่ โรงเรียนที่มีผู้บริหารที่เล็งเห็นความสำคัญและมุ่งเน้นการพัฒนางานวิชาการ มีความเข้าใจเกี่ยวกับการปฏิรูปการเรียนรู้ สามารถแปลงความคิดนำไปสู่การปฏิบัติอีกทั้งยังต้องเป็นแบบอย่างที่ดีในการปฏิบัติงาน นอกจากปัจจัยภายใน ซึ่งได้แก่ โครงสร้างองค์กร บุคลากร สภาพทางกายภาพ งบประมาณ และการบริหารจัดการที่เกี่ยวข้องกับคุณภาพการเรียนรู้ของนักเรียน ยังมีปัจจัยภายนอกอีกหลายประการซึ่งในอนาคตปัจจัยดังกล่าวอาจจะส่งผลกระทบต่อระบบการจัดการศึกษาของประเทศได้เช่นเดียวกัน เช่น ปัจจัยแรก คือ ด้านเทคโนโลยี การจัดการศึกษาต้องมีการเพิ่มพูนความรู้เกี่ยวกับเทคโนโลยีที่ทันสมัยในหลักสูตรการเรียนการสอนและให้ทันต่อการเปลี่ยนแปลงของเทคโนโลยีที่เปลี่ยนแปลงตลอดเวลา ปัจจัยที่สอง คือ ด้านเศรษฐกิจซึ่งมีผลต่อตลาดแรงงานและตลาดการศึกษาจำเป็นที่ต้องต้องพัฒนาบุคลากรให้มีทักษะการทำวิจัย ปัจจัยที่สาม คือ ด้านระบบราชการซึ่งที่ผ่านมาปฏิรูปการศึกษาเคลื่อนไปอย่างยากลำบากเพราะระบบราชการเป็นปัญหาและมีอุปสรรคที่ไม่สนับสนุน การปฏิรูปการศึกษา ปัจจัยที่สี่ คือ ด้านการเมืองโดยการปฏิรูปหรือพัฒนาการจัดการศึกษาจะสำเร็จหรือล้มเหลวขึ้นอยู่กับการเมืองมากกว่าแนวทางและวิธีการปฏิบัติอื่นๆ (อภิเชษฐ์ ฉิมพลีสวรรค์, 2552)

4.4.2 การใช้โมเดลมูลค่าเพิ่มเพื่อประเมินคุณภาพการจัดการศึกษา

งานวิจัยและเอกสารที่เกี่ยวข้องกับการศึกษาการใช้โมเดลมูลค่าเพิ่มเพื่อประเมินคุณภาพ การจัดการศึกษาทั้งในประเทศและต่างประเทศมีหลากหลายด้าน ผู้วิจัยจำแนกออกเป็น 4 กลุ่ม ได้แก่ กลุ่มที่ 1 ศึกษาการสร้างโมเดลโดยการวิเคราะห์มูลค่าเพิ่ม (สุบิน ยุระรัช, 2547;

จตุภูมิ เขตจัตุรัส, 2552; Tekwe et al., 2004) กลุ่มที่ 2 ศึกษาปัจจัยและความสัมพันธ์โดยการวิเคราะห์มูลค่าเพิ่ม (สุชีรา มะหิเมือง, 2547; Keeves, Hungi, & Afrassa, 2005; Noell & Burns, 2008; Ready, 2013) กลุ่มที่ 3 ศึกษาประยุกต์ใช้การวิเคราะห์มูลค่าเพิ่ม (Tuksino, Kanjanawasee, & Pasiphol, 2010; วัยวุฒฒ์ อยู่ในศิล,2553; Sirikit, Pasiphol, & Kanjanawasee, 2012) และกลุ่มที่ 4 ศึกษาพัฒนาองค์ความรู้ในการวิเคราะห์มูลค่าเพิ่ม (Steedle, 2010)

กลุ่มที่ศึกษาโดยการสร้างโมเดลโดยการวิเคราะห์มูลค่าเพิ่ม ได้แก่ สุบิน ยุระรัช (2547) ได้สร้างโมเดลสามระดับของข้อมูลตัวแปรที่ส่งผลต่อคุณภาพนักเรียนประถมศึกษาในกรุงเทพมหานคร โดยใช้โปรแกรมเอ็มแอลวิน พบว่า การสร้างโมเดลการวิเคราะห์ข้อมูลตัวแปร 3 ระดับ ที่ส่งผลต่อคุณภาพนักเรียนประถมศึกษาในกรุงเทพมหานคร สร้างได้เฉพาะโมเดลระดับนักเรียน (ระดับที่ 1) ซึ่งผลการวิเคราะห์โมเดลสามระดับที่ให้ค่าคงที่ผันแปรได้ (three-level random intercept model) พบว่า ในโมเดลระดับนักเรียนที่สร้างได้ มีตัวแปรทำนายระดับนักเรียน 5 ตัวแปรที่ส่งผลต่อคุณภาพนักเรียนประถมศึกษาอย่างมีนัยสำคัญทางสถิติ ได้แก่ ความคาดหวังของนักเรียนในการศึกษาต่อ ความคิดเห็นของนักเรียนที่มีต่อการเรียน พฤติกรรมการเรียนของนักเรียน แรงจูงใจใฝ่สัมฤทธิ์ของนักเรียน และพฤติกรรมของบิดามารดาในการส่งเสริมการเรียนของบุตร โมเดลระดับห้องเรียน พบว่า ไม่มีตัวแปรใดส่งผลต่อค่าเฉลี่ยคุณภาพนักเรียนประถมศึกษา ส่วนโมเดลระดับโรงเรียนไม่นำมาวิเคราะห์เนื่องจากไม่พบความผันแปรของคุณภาพนักเรียนประถมศึกษาในระดับโรงเรียน

ต่อมาจตุภูมิ เขตจัตุรัส (2552) ได้พัฒนาโมเดลมูลค่าเพิ่มของผลสัมฤทธิ์ทางวิชาการและแบบตรวจสอบรายการประเมินตนเองเพื่อเพิ่มมูลค่ากระบวนการจัดการศึกษา พบว่า โมเดลมูลค่าเพิ่มของผลสัมฤทธิ์ทางวิชาการ มีลักษณะเป็นโมเดลพัฒนาการเชิงเส้นแบบพหุตัวแปรที่มีอิทธิพลสุ่มแบบลดหลั่น 3 ระดับ โดยใช้ข้อมูลคะแนนเฉลี่ยผลสัมฤทธิ์ทางวิชาการของนักเรียนระดับมัธยมศึกษาปีที่ 6 ในแต่ละปีการศึกษาของสถานศึกษาที่ตั้งอยู่สำนักงานเขตพื้นที่การศึกษา โมเดลสามารถอธิบายความแปรปรวนขอตัวแปรตามทั้งหมดได้ร้อยละ 58-82 ผลการวิเคราะห์โมเดล พบว่า ตัวแปรระดับสถานศึกษาที่มีความสำคัญต่อผลสัมฤทธิ์และความก้าวหน้าทางวิชาการ ได้แก่ ขนาดสถานศึกษา การจัดกิจกรรมและการเรียนการสอนโดยเน้นผู้เรียนเป็นสำคัญ การพัฒนาหลักสูตรและสื่อที่เอื้อต่อการเรียนรู้ของผู้เรียนและท้องถิ่น และตัวแปรระดับเขตพื้นที่ ได้แก่ เศรษฐานะเขตพื้นที่ และการยกระดับผลสัมฤทธิ์ทางการเรียน การเพิ่มมูลค่ากระบวนการจัดการศึกษา มี 3 แบบ ได้แก่ การเพิ่มขีดความสามารถในการพัฒนาคุณภาพการจัดการศึกษา การเพิ่มขีดความสามารถในการตอบสนองความต้องการของผู้เรียนและการเพิ่มขีดความสามารถในการปรับตัวให้เข้ากับสิ่งแวดล้อม โดยสอดคล้องกับแนวคิดการเพิ่มมูลค่า 5 ลักษณะ ได้แก่ การใช้ความรู้อย่างเข้มข้น การเพิ่มผลผลิต การสร้างความพึงพอใจ การแก้ไขและพัฒนาคุณลักษณะที่พึง

ประสงค์ และการวิเคราะห์พัฒนาการ/ความก้าวหน้า และแบบตรวจสอบรายการประเมินตนเองเพื่อเพิ่มมูลค่ากระบวนการจัดการศึกษา ประกอบด้วย 3 จุดตรวจสอบหลัก 14 จุดตรวจสอบย่อย มีค่า IOC ของข้อรายการย่อย ระหว่าง 0.538-1.000 และค่าความเที่ยงทั้งฉบับ ระหว่าง 0.955-0.9888 กระบวนการนำแบบตรวจสอบรายการประเมินตนเองไปใช้มีลักษณะเป็นวงจร PDCA ความคิดเห็นของผู้ใช้ต่อคุณภาพของแบบตรวจสอบรายการประเมินตนเองใน 4 ด้าน ได้แก่ ความถูกต้องครอบคลุม ความมีประโยชน์ ความเป็นไปได้ในการนำไปใช้ และความเหมาะสมอยู่ในระดับมากทุกด้าน (Mean=3.88, 4.13, 3.98, 4.06 ตามลำดับ) ผลของการใช้แบบตรวจสอบรายการประเมินตนเองช่วยทำให้สถานศึกษาและสำนักงานเขตพื้นที่การศึกษารู้จุดที่ต้องทำการเพิ่มมูลค่า และสามารถกำหนดทิศทางการพัฒนางานได้

กลุ่มที่ศึกษาโดยการสร้างโมเดลโดยการวิเคราะห์มูลค่าเพิ่มนอกจาก สุบิน ยุระรัช (2547) และ จตุภูมิ เขตจัตุรัส (2552) ยังมี Tekwe และคณะ (2004) ซึ่งได้เปรียบเทียบโมเดลทางสถิติสำหรับประเมินมูลค่าเพิ่มของการปฏิบัติงานโรงเรียน โดยใช้ simple fixed effects model (SFEM) และ layered mixed effects model (LMEM) ซึ่ง LMEM เป็นการวิเคราะห์พัฒนาการเรียนรู้โดยไม่พิจารณาปัจจัยทางด้านคุณลักษณะประชาการ เช่น ความสามารถในตนต่อพัฒนาการเรียนรู้ระหว่างหลายโรงเรียนและอัตถประโยชน์ของการสำรวจที่ไม่สมบูรณ์ จากการวิจัยพบว่า SFEM มีความสัมพันธ์สูงกับ LMEM แต่ควรใช้วิธี SFEM ในการวิเคราะห์มูลค่าเพิ่มมากกว่า LMEM

กลุ่มที่ศึกษาปัจจัยและความสัมพันธ์โดยการวิเคราะห์มูลค่าเพิ่ม ได้แก่ สุชีรา มะหิเมือง (2547) ซึ่งศึกษาปัจจัยที่ส่งผลต่อผลสัมฤทธิ์และพัฒนาการทางวิชาการ โดยการวิเคราะห์มูลค่าเพิ่มพบว่า โมเดลการวิเคราะห์มูลค่าเพิ่มของค่าเฉลี่ยผลสัมฤทธิ์และพัฒนาการทางวิชาการที่ใช้วิธีวิเคราะห์ถดถอยแบบพหุระดับด้วยโปรแกรม HLM ซึ่งกำหนดให้ตัวแปรภูมิหลังของนักเรียนด้านผลสัมฤทธิ์เดิมเป็นตัวแปรที่ส่งผลเชิงสุ่ม ในขณะที่ควบคุมอิทธิพลจากตัวแปรด้านเศรษฐานะ และพื้นฐานทางภาษา เป็นโมเดลที่มีประสิทธิภาพในการพยากรณ์ (R^2) 22.10% และ 24.97% ในวิชาภาษาไทย และคณิตศาสตร์ ตามลำดับ การแจกแจงความถี่ของมูลค่าเพิ่มของค่าเฉลี่ยผลสัมฤทธิ์และพัฒนาการวิชาการทั้งสองวิชาโดยรวมมีลักษณะเป็นโค้งปกติด้วยค่าเฉลี่ยที่ใกล้เคียงหรือเท่ากับ ‘0’ แต่มีความแปรปรวนของมูลค่าเพิ่มของค่าเฉลี่ยผลสัมฤทธิ์วิชาภาษาไทยและคณิตศาสตร์ (ระดับนัยสำคัญ .05 และ .01 ตามลำดับ) ระหว่างจังหวัดตากกับจังหวัดสระบุรี และของมูลค่าเพิ่มของพัฒนาการทางวิชาการคณิตศาสตร์ (ระดับนัยสำคัญ .01) ระหว่างจังหวัดตากและจังหวัดระนอง นอกจากนี้โรงเรียนที่มีมูลค่าเพิ่มฯ จัดอยู่ในกลุ่มสูงเป็นโรงเรียนที่มีบริบททั่วไปของโรงเรียนและสภาพการปฏิบัติงานทางวิชาการดีกว่ากลุ่มโรงเรียนที่มีมูลค่าเพิ่มฯ ในระดับต่ำ และโมเดลเชิงสาเหตุแสดงอิทธิพลของปัจจัย/ตัวแปรระดับโรงเรียนที่มีต่อมูลค่าเพิ่มฯ ซึ่งกำหนดขึ้น

ตามกรอบความคิดเชิงทฤษฎี เป็นโมเดลที่มีความตรงและความกลมกลืนกับข้อมูลเชิงประจักษ์ ตัวแปรที่มีอิทธิพลอย่างมีนัยสำคัญ (ระดับ .05) ได้แก่ การปฏิบัติงานทางวิชาการที่ส่งผลเชิงบวกต่อมูลค่าเพิ่มของ ค่าเฉลี่ยผลสัมฤทธิ์ และเชิงลบต่อมูลค่าเพิ่มของพัฒนาการทางวิชาการ ทำเลที่ตั้งของโรงเรียนที่ส่งผลเชิงลบต่อมูลค่าเพิ่มของพัฒนาการทางวิชาการวิชาคณิตศาสตร์ และการส่งผลเชิงบวกของมูลค่าเพิ่มของค่าเฉลี่ย ผลสัมฤทธิ์ต่อมูลค่าเพิ่มของพัฒนาการทางวิชาการ (ระดับนัยสำคัญ .01)

นอกจาก สุชีรา มะหิเมือง (2547) แล้ว Keeves, Hungi, และ Afrassa (2005) ได้ศึกษาการวัดผลมูลค่าเพิ่มข้ามโรงเรียนในแบบทดสอบทักษะพื้นฐานการอ่านและตัวเลข โดยทดสอบกับนักเรียน 2 ส่วน ส่วนแรกมี 2 ระดับคือ เกรด 3 และเกรด 5 และส่วนที่สองมี 3 ระดับคือ เกรด 3 เกรด 5 และ เกรด 7 แบบทดสอบถูกทดสอบโดยใช้วิธี rasch scaling และการปรับเทียบโดยใช้กระบวนการปรับเทียบที่เป็นปัจจุบัน (concurrent equating procedures) กับนักเรียนประมาณ 8,000 คน จาก 440 โรงเรียน และใช้การวิเคราะห์โมเดลลดหลั่นเพื่อวิเคราะห์การเปลี่ยนแปลงเชิงสัมพัทธ์และเชิงสัมบูรณ์ของความสามารถ จากการวิจัย พบว่า โมเดลถดถอยที่ต่างกันและตัวแปรที่ต่างกันในการพิจารณาตัวประมาณค่าของความแปรปรวนระหว่างโรงเรียน ในบางสถานการณ์ความแปรปรวนเศษเหลือระหว่างโรงเรียนมีขนาดน้อยมาก

นอกจากผู้ที่ศึกษาปัจจัยและความสัมพันธ์โดยการวิเคราะห์มูลค่าเพิ่ม คือ สุชีรา มะหิเมือง (2547) และ Keeves, Hungi, และ Afrassa (2005) ยังมี Noell และ Burns (2008) ซึ่งวัดมูลค่าเพิ่มของการเตรียมการเป็นครูในรัฐลุสเซียน่า ระหว่างปี 2004-2005 ถึง 2006-2007 โดยใช้การวิเคราะห์ลดหลั่นแบบซ้อนทับ (nested) ซึ่งนักเรียนอยู่ภายในอาจารย์ และอาจารย์อยู่ภายในโรงเรียน โมเดลที่ใช้สำหรับใช้ในการทำนายผลสัมฤทธิ์ของนักเรียนที่ขึ้นอยู่กับผลสัมฤทธิ์เบื้องต้น ปัจจัยด้านจำนวนนักเรียนและระดับชั้นเรียนร่วม จากการวิจัยพบว่า คะแนน ACT วิชาคณิตศาสตร์เป็นตัวทำนายทางสถิติที่ดีในด้านประสิทธิผลความเป็นครูคณิตศาสตร์ ส่วนครูที่ไม่มีใบรับประกอบวิชาชีพในวิชาที่สอนมีประสิทธิผลน้อยกว่าครูที่มีใบประกอบวิชาชีพ

นอกจากนี้ Ready (2013) ยังศึกษาความสัมพันธ์ระหว่างผลสัมฤทธิ์ของนักเรียนและการเรียนรู้โดยใช้โมเดลการตรวจสอบมูลค่าเพิ่ม พบว่า โมเดลมูลค่าเพิ่มเป็นระบบการตรวจสอบที่ใช้วัดการเรียนรู้ของนักเรียนมากกว่าผลสัมฤทธิ์ของนักเรียนซึ่งมีศักยภาพต่อการประเมินคุณภาพโรงเรียน การศึกษาครั้งนี้ใช้ข้อมูลการวัดผลของรัฐระดับนักเรียนเพื่อสำรวจความสัมพันธ์สำหรับตัวบ่งชี้การวัดมูลเพิ่ม พบว่า ความสัมพันธ์ระหว่างผลสัมฤทธิ์กับพัฒนาผลสัมฤทธิ์ระดับย่อยมีความสัมพันธ์ทางลบสูงและความสัมพันธ์ระดับนักเรียนมีอิทธิพลอย่างมากต่อการประเมินมูลค่าเพิ่มระดับโรงเรียน

กลุ่มที่ศึกษาประยุกต์ใช้การวิเคราะห์มูลค่าเพิ่ม ได้แก่ Tuksino, Kanjanawasee, และ Pasiphol (2010) ซึ่งประเมินคุณภาพการจัดการศึกษาวิชาวิทยาศาสตร์ของสถานศึกษาขั้นพื้นฐาน โดยเป็นการประยุกต์ใช้การทำหน้าที่ต่างกันของข้อสอบและโมเดลมูลค่าเพิ่ม (value-added model) ประกอบด้วย 4 โมเดลที่ต่างกัน โดยโมเดล 1 ไม่ตัดข้อสอบที่ทำหน้าที่ต่างกันออกและไม่มีการปรับแก้คะแนนด้วยตัวแปรนักเรียนและสถานศึกษา โมเดล 2 ไม่ตัดข้อสอบที่ทำหน้าที่ต่างกันออกและมีการปรับแก้คะแนนด้วยตัวแปรนักเรียนและสถานศึกษา โมเดล 3 ตัดข้อสอบที่ทำหน้าที่ต่างกันออกและไม่มีการปรับแก้คะแนนด้วยตัวแปรนักเรียนและสถานศึกษา โมเดล 4 ตัดข้อสอบที่ทำหน้าที่ต่างกันออกและมีการปรับแก้คะแนนด้วยตัวแปรนักเรียนและสถานศึกษา ซึ่งผลของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบวิชาวิทยาศาสตร์ต่อการประเมินคุณภาพการจัดการศึกษาวิชาวิทยาศาสตร์ของสถานศึกษาขั้นพื้นฐาน พบว่า ข้อสอบที่ทำหน้าที่ต่างกันในแบบวิชาวิทยาศาสตร์ไม่มีผลต่อการประเมินคุณภาพการจัดการศึกษา (โมเดล 1 และ 3) และเมื่อมีการควบคุมอิทธิพลของคุณลักษณะของนักเรียนและสถานศึกษา (โมเดล 2 และ 4) พบว่า ข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบวิชาวิทยาศาสตร์ทำให้การจัดระดับคุณภาพการจัดการศึกษาของสถานศึกษาแตกต่างกันอย่างมีนัยสำคัญที่ระดับ .05 และผลของตัวแปรคุณลักษณะของนักเรียนและสถานศึกษามีผลต่อการประเมินคุณภาพการจัดการศึกษาของสถานศึกษาโดยการจัดกลุ่มคุณภาพและระดับคุณภาพการจัดการศึกษาของสถานศึกษาระหว่างโมเดล 1 และ 2 และระหว่างโมเดล 3 และ 4 มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และ โมเดล 4 มีสัมประสิทธิ์การทำนายสูงสุดเท่ากับ 88.63% และโมเดล 2 เท่ากับ 77.75% และสหสัมพันธ์ตำแหน่งคะแนนของสเปียร์แมนของโมเดล 2 และ 4 เท่ากับ 100 และโมเดล 1 กับ 3 เท่ากับ 0.981

ต่อมา วัยวุฒฒ์ อยู่ในศิล (2553) ได้ประยุกต์ใช้การวิเคราะห์สายโซ่วิธีการเป้าหมายในการพัฒนาคุณภาพการเรียนการสอนที่เน้นผู้เรียนเป็นศูนย์กลางโดยการพัฒนาโมเดลการวัดและการวิเคราะห์มูลค่าเพิ่ม พบว่า โมเดลการวัดคุณภาพของการเรียนการสอนที่เน้นผู้เรียนเป็นศูนย์กลางตามแนวการวิเคราะห์สายโซ่วิธีการ-เป้าหมายเมื่อกำหนดคุณภาพการเรียนการสอน วัดได้จากตัวบ่งชี้ 3 ตัวคือคุณลักษณะ ผลที่เกิดตามมา และคุณค่าซึ่งมีอิทธิพลต่อตัวบ่งชี้ที่ตามมา มีความสอดคล้องกับข้อมูลเชิงประจักษ์ มีค่า Chi-square = 40.977, df=29, P=0.069, RMSEA= 0.020, RMR= 0.003, GFI= 0.993, AGFI=0.982 และ CFI=.999 ผลการวิเคราะห์พบว่า ในจำนวนตัวบ่งชี้ 3 ตัวมีเพียงตัวบ่งชี้ที่เป็นอิทธิพลทางตรงและทางอ้อมของคุณลักษณะต่อคุณค่าผ่านผลที่เกิดตามมา การเปรียบเทียบผลการวัดคุณภาพการเรียนการสอนระหว่างเครื่องมือวัดทั้งสองแบบ พบว่า ผลการวัดมีความสอดคล้องกันน้อย ส่วนมูลค่าเพิ่มตามแนวคิดการวิเคราะห์สายโซ่วิธีการ-เป้าหมาย ประกอบด้วย ผู้เรียนมีความกล้าในการใช้ภาษา มีโอกาสในการศึกษาต่อในระดับสูง มีความภูมิใจและเห็นคุณค่าในการศึกษาวัฒนธรรม เกิดใจกว้างยอมรับวัฒนธรรม

มีแรงจูงใจในการปฏิบัติตนตามหน้าที่ และมีความเชื่อมั่นในบุคลิกภาพของตนเอง เพิ่มขึ้นจากการได้เรียนกับครูชาวต่างและประเทศ และปัจจัยที่ส่งผลต่อคุณภาพการเรียนการสอนที่เน้นผู้เรียนเป็นศูนย์กลาง ตามแนวการวิเคราะห์สายโซ่วิธีการเป้าหมาย พบว่า ตัวแปรเพศ เกรดเฉลี่ย เกรดเฉลี่ยวิชาภาษาอังกฤษ และประสบการณ์การเรียนภาษาอังกฤษกับครูชาวต่างประเทศส่งอิทธิพลทั้งทางตรงและทางอ้อมที่ได้รับผ่านทางตัวแปรคุณลักษณะและผลที่เกิดตามมา โดยคุณค่าได้รับอิทธิพลทางอ้อมจากตัวแปรอิสระทั้ง 4 ตัวสูงกว่าอิทธิพลทางตรง เกรดเฉลี่ยวิชาภาษาอังกฤษในภาคเรียนที่ผ่านมาส่งอิทธิพลทางอ้อมต่อคุณค่าด้วยขนาดอิทธิพลสูงสุด

นอกจากผู้ที่ศึกษาประยุกต์ใช้การวิเคราะห์มูลค่าเพิ่ม ได้แก่ Tuksino, Kanjanawasee, และ Pasiphol (2010) และวัยวุฒฑ์ อยู่ในศีล (2553) ต่อมา Sirikit, Pasiphol, และ Kanjanawasee (2012) ซึ่งได้วิเคราะห์เปรียบเทียบโมเดลการประเมินคุณภาพการจัดการศึกษาวิชาคณิตศาสตร์ 4 โมเดลที่ต่างกันกัน คือ โมเดล 1 Undetected DIF-DDF & Adjusted โมเดล 2 Detect DIF & Adjusted โมเดล 3 Detect DIF-DDF & Unadjusted และโมเดล 4 Detected DIF-DDF & Adjusted โดยประยุกต์ใช้โมเดลมูลค่าเพิ่มที่มีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและการวิเคราะห์การทำหน้าที่ต่างกันของตัวลวง พบว่า ตัวแปรเพศเป็นคุณลักษณะที่พบการทำหน้าที่ต่างกันของข้อสอบ และการทำหน้าที่ต่างกันของตัวลวงในแบบสอบวิชาคณิตศาสตร์มากที่สุด การทำหน้าที่ต่างกันของจำนวนข้อสอบที่เพศหญิงมีโอกาสที่จะตอบถูกมากกว่าในเพศชาย ส่วนการทำหน้าที่ต่างกันของตัวลวงข้อสอบส่วนใหญ่เพศชายมีโอกาสที่จะเลือกตอบมากกว่าเพศหญิง ผลของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบวิชาคณิตศาสตร์ต่อการประเมินคุณภาพการจัดการศึกษา พบว่า ข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบวิชาคณิตศาสตร์ไม่มีผลต่อการประเมินคุณภาพการจัดการศึกษา ส่วนผลของตัวแปรคุณลักษณะของนักเรียนและสถานศึกษาต่อการประเมินคุณภาพการจัดการศึกษาวิชาคณิตศาสตร์ พบว่าคุณลักษณะของนักเรียนและสถานศึกษามีผลต่อการประเมินคุณภาพการจัดการศึกษาของศึกษา

และกลุ่มที่ศึกษาพัฒนาองค์ความรู้ในการวิเคราะห์มูลค่าเพิ่ม ได้แก่ Steedle (2010) ซึ่งพัฒนาความเที่ยงและความสามารถในการแปลผลของคะแนนมูลค่าเพิ่มสำหรับโปรแกรมการประเมินมัธยมศึกษาตอนปลาย โดยเป็นการศึกษาเปรียบเทียบเชิงปริมาณของคะแนนมูลค่าเพิ่มจากแนวคิดมูลค่าเพิ่มการวัดการเรียนรู้เกี่ยวกับสถานศึกษาเบื้องต้น (original ccollegiate learning assessment value-added approach) และโมเดลเชิงลดหลั่น (HLM) ปรากฏว่า คะแนนมูลค่าเพิ่มจากสองแนวคิดมีความสัมพันธ์กัน ซึ่งวิธี HLM ให้รายละเอียดในการแปลผลจากการระบุตัวบ่งชี้ของโรงเรียนผ่านคะแนนในโมเดลมูลค่าเพิ่ม

จากการศึกษางานวิจัยที่เกี่ยวข้องกับการใช้โมเดลมูลค่าเพิ่มเพื่อประเมินคุณภาพการจัดการศึกษา สามารถสรุปผลการวิเคราะห์องค์ความรู้ที่ได้รับรายละเอียดตามตารางที่ 13 ดังนี้

ตารางที่ 13 สรุปผลองค์ความรู้จากงานวิจัยโมเดลมูลค่าเพิ่ม

ผู้วิจัย	วิธีการดำเนินงาน	องค์ความรู้ในการศึกษา
สุบิน ยุระรัช (2547)	สร้างโมเดลสามระดับของข้อมูลตัวแปรที่ส่งผลต่อคุณภาพนักเรียนประถมศึกษาในกรุงเทพมหานครโดยใช้โปรแกรมเอ็มแอลวิน	วิเคราะห์โมเดลสามระดับที่ให้ค่าคงที่ผันแปรได้ (three-level random intercept model)
สุชีรา มะหิเมือง (2547)	ศึกษาปัจจัยที่ส่งผลต่อผลสัมฤทธิ์และพัฒนาการทางวิชาการ โดยการวิเคราะห์มูลค่าเพิ่ม	วิเคราะห์มูลค่าเพิ่มของค่าเฉลี่ยผลสัมฤทธิ์และพัฒนาการทางวิชาการที่ใช้วิธีวิเคราะห์ถดถอยแบบพหุระดับด้วยโปรแกรม HLM
จตุภูมิ เขตจัตุรัส (2552)	พัฒนาโมเดลมูลค่าเพิ่มของผลสัมฤทธิ์ทางวิชาการและแบบตรวจสอบรายการประเมินตนเองเพื่อเพิ่มมูลค่ากระบวนการจัดการศึกษา	โมเดลมูลค่าเพิ่มของผลสัมฤทธิ์ทางวิชาการ มีลักษณะเป็นโมเดลพัฒนาการเชิงเส้นแบบพหุตัวแปรที่มีอิทธิพลสุ่มแบบลดหลั่น 3 ระดับ
Tuksino, Kanjanawasee, และ Pasiphol (2010)	ประเมินคุณภาพการจัดการศึกษาวิชาวิทยาศาสตร์	ประยุกต์ใช้การทำหน้าที่ต่างกันของข้อสอบและโมเดลมูลค่าเพิ่ม
วัยวุฒฒ์ อยู่ในศิล (2553)	ประยุกต์ใช้การวิเคราะห์สายโซ่วิธีการ-เป้าหมายในการพัฒนาคุณภาพการเรียนการสอนที่เน้นผู้เรียนเป็นศูนย์กลางโดยการพัฒนาโมเดลการวัดและการวิเคราะห์มูลค่าเพิ่ม	วิเคราะห์มูลค่าเพิ่มตามแนวคิดการวิเคราะห์สายโซ่วิธีการ-เป้าหมาย
Sirikit, Pasiphol, และ Kanjanawasee (2012)	เปรียบเทียบโมเดลการประเมินคุณภาพการจัดการศึกษา	ประยุกต์ใช้โมเดลมูลค่าเพิ่มที่มีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและการทำหน้าที่ต่างกันของตัวลวง
Tekwe และคณะ (2004)	เปรียบเทียบเชิงประจักษ์ของโมเดลทางสถิติสำหรับประเมินมูลค่าเพิ่มของการปฏิบัติงานโรงเรียน	ประเมินมูลค่าเพิ่มของการปฏิบัติงานโรงเรียน โดยใช้ simple fixed effects model (SFEM) และ layered mixed effects model (LMEM)

ตารางที่ 13 สรุปผลองค์ความรู้จากงานวิจัยโมเดลมูลค่าเพิ่ม (ต่อ)

ผู้วิจัย	วิธีการดำเนินงาน	องค์ความรู้ในการศึกษา
Keeves, Hungi, และ Afrassa (2005)	วัดมูลค่าเพิ่มข้ามโรงเรียน	ใช้โมเดลลดหลั่นวิเคราะห์การเปลี่ยนเชิงสัมพัทธ์และเชิงสัมบูรณ์ของความสามารถ
Noell และBurns (2008)	วัดมูลค่าเพิ่มการเตรียมการเป็นครู	ใช้การวัดมูลค่าเพิ่มของการเตรียมการและทำนายประสิทธิผลความเป็นครู
Steedle (2010)	พัฒนาความเที่ยงและความสามารถในการแปลผลของคะแนนมูลค่าเพิ่มสำหรับโปรแกรมการประเมินมัธยมศึกษาตอนปลาย	ใช้คะแนนมูลค่าเพิ่มมาพัฒนาความสามารถในการแปลผลการประเมิน
Ready (2013)	ศึกษาความสัมพันธ์ระหว่างผลสัมฤทธิ์ของนักเรียนและการเรียนรู้โดยใช้โมเดลการตรวจสอบได้มูลค่าเพิ่ม	ใช้โมเดลการตรวจสอบได้มูลค่าเพิ่มมาศึกษาผลสัมฤทธิ์และการเรียนรู้ของนักเรียน

มูลค่าเพิ่ม (value-added) เป็นแนวคิดที่ได้มีการพัฒนามาจากเศรษฐศาสตร์ ถูกใช้เป็นเครื่องมือวัดผลการดำเนินงานทางธุรกิจในเชิงเศรษฐศาสตร์ โดยคำนึงถึงผลต่างระหว่างมูลค่าของผลผลิต (output) กับมูลค่าของปัจจัยนำเข้า (input) แนวคิดดังกล่าวได้ถูกนำมาใช้ทางการศึกษาเพื่ออธิบายการเพิ่มคุณค่า (value) ของสถานศึกษาสามารถทำนายจากพื้นฐานของนักเรียนและความสำเร็จเดิมของนักเรียน เช่น เพศ เชื้อชาติ เศรษฐานะ ผลสัมฤทธิ์ทางการเรียนเดิม วิธีการวัดมูลค่าเพิ่มได้มีการนำมาใช้หลายประเทศเพื่อใช้ผลมาพิจารณาการดำเนินการของสถานศึกษาที่ส่งผลต่อผลสำเร็จของนักเรียน ดังตัวอย่างของระบบการประเมินในรัฐเทนเนสซีของประเทศสหรัฐอเมริกาที่มีชื่อระบบการประเมินว่าระบบการวัดมูลค่าเพิ่มของเทนเนสซี (tennessess value-added assessment system: TVAAS) ซึ่งพัฒนาโดย William L. Sanders และมีการศึกษาเพิ่มเติมไปสู่การดำเนินการระบบการประเมินมูลค่าเพิ่มทางการศึกษา (education value-added assessment system: EVAAS) นอกจากนี้ผลการประเมินมูลค่าเพิ่มมีความสำคัญกับการนำมาใช้กำหนดนโยบายที่สำคัญ 3 ด้าน คือ การริเริ่มการพัฒนาสถานศึกษา (school improvement initiatives) ระบบความสามารถในการตรวจสอบได้ของสถานศึกษา (systems of school accountability) และการเลือกสถานศึกษา (school choice) ทั้งนี้ข้อมูลที่ได้จากการวิเคราะห์มูลค่าเพิ่มมีบทบาทสำคัญต่อนโยบาย การจัดสรรแหล่งทรัพยากร และกลยุทธ์การสอนที่จำเป็นในแต่ละสถานศึกษา การวัดมูลค่าเพิ่มมีความสำคัญหลายประการสำหรับผู้ที่เกี่ยวข้องหลายฝ่าย เช่น

การประเมินนโยบายและโครงการ (policy and program evaluation) เครื่องมือหนึ่งสำหรับการพัฒนาสถานศึกษา (a tool for school improvement) วัดประสิทธิผลของครู (teacher effectiveness) เป็นเครื่องมือหนึ่งสำหรับความรับผิดชอบที่ตรวจสอบได้ (a tool for accountability) การรายงานต่อผู้ปกครองและชุมชน (reporting to parents and community)

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับปัจจัยหรือตัวแปรที่ส่งผลต่อคุณภาพการจัดการศึกษาและการใช้โมเดลมูลค่าเพิ่มเพื่อประเมินคุณภาพการจัดการศึกษา มีการศึกษาจากนักวิชาการที่หลากหลายในทั้งประเทศและต่างประเทศ ในส่วนงานวิจัยและเอกสารที่เกี่ยวข้องกับปัจจัยหรือตัวแปรที่สงผลต่อคุณภาพการจัดการศึกษา จำแนกออกได้เป็น 2 ส่วนหลัก ได้แก่ ระดับนักเรียนซึ่งเป็นเป้าหมายหลักที่ต้องการให้มีคุณภาพ ส่วนระดับสถานศึกษาซึ่งเป็นสถาบันรองรับนักเรียนเกี่ยวข้องกับการบริหารจัดการที่ส่งผลต่อคุณภาพของนักเรียน โดยการสนับสนุนในเรื่องสิ่งอำนวยความสะดวกต่อการเรียนรู้ของนักเรียน ส่วนการใช้โมเดลมูลค่าเพิ่มเพื่อประเมินคุณภาพการจัดการศึกษา สามารถสังเคราะห์งานวิจัยออกเป็น 4 กลุ่ม ได้แก่ กลุ่มการสร้างโมเดลโดยการวิเคราะห์มูลค่าเพิ่ม กลุ่มปัจจัยและความสัมพันธ์โดยการวิเคราะห์มูลค่าเพิ่ม กลุ่มการประยุกต์ใช้การวิเคราะห์มูลค่าเพิ่มและกลุ่มการพัฒนาองค์ความรู้ในการวิเคราะห์มูลค่าเพิ่ม

อย่างไรก็ตาม งานวิจัยส่วนใหญ่ยังเป็นการศึกษาปัจจัยและความสัมพันธ์โดยอาศัยการวิเคราะห์มูลค่าเพิ่ม ส่วนการประยุกต์ใช้โมเดลมูลค่าเพิ่มทางด้านการวัดผลและประเมินผลมีเฉพาะการวิจัย ของจตุภูมิ เขตจัตุรัส (2552) Tuksino, Kanjanawasee, และ Pasiphol (2010) และ Sirikit, Pasiphol, และ Kanjanawasee (2012) ซึ่งจำนวนงานวิจัยในเชิงการวัดผลและประเมินผลมีค่อนข้างน้อย จากเหตุผลข้างต้นและเอกสารและงานวิจัยที่เกี่ยวข้อง ผู้วิจัยจึงมีความสนใจที่จะนำองค์ความรู้ทางด้านมูลค่าเพิ่มการจัดการศึกษาเข้ามาศึกษาในบริบททางการวัดผลและประเมินผลในประเทศไทย นอกจากนี้ผู้วิจัยยังได้สังเคราะห์เอกสารและงานวิจัยที่เกี่ยวข้องเพื่อนำมาใช้กำหนดแนวคิดในการวิจัย จากผลการสังเคราะห์งานวิจัยจำนวน 12 เรื่อง และจากฐานข้อมูลของ PISA 2009 รายละเอียดผลการสังเคราะห์ดังตารางที่ 14

ตารางที่ 14 ผลการสังเคราะห์ตัวแปรทำนายที่มีอิทธิพลต่อผลสัมฤทธิ์ทางการเรียน

ตัวแปร	สุบิน ยุระวัช (2547)	สุชีรา มะหิเมือง (2547)	จตุภูมิ เขตจัตุรัส (2552)	Tuksino, Kanjanawasee, และ Pasiphol (2010)	วัยวุฒิ อยู่ในศิล (2553)	Sirikit, Pasiphol, และ Kanjanawasee (2012)	Tekwe และคณะ (2004)	Keeves, Hungi, และ Afrassa (2005)	Noell และ Burns (2008)	Steedle (2010)	Ready (2013)	PISSA (2009)	รวม
ระดับสถานศึกษา													
ขนาดสถานศึกษา	✓	✓	✓	✓		✓		✓				✓	7
ที่ตั้งสถานศึกษา	✓	✓		✓		✓		✓				✓	6
บริบทสถานศึกษา/สังกัด				✓								✓	2
งบประมาณ/รายได้ของหน่วยงาน		✓	✓	✓			✓					✓	5
แหล่งทรัพยากร อุปกรณ์				✓		✓					✓	✓	4
การขาดแคลนครู	✓		✓	✓		✓						✓	5
ประสบการณ์สอนของครู				✓		✓						✓	3
ความสามารถในการจัดการเรียนการสอน	✓	✓	✓			✓							4
หลักสูตรและการประเมินผล		✓	✓	✓								✓	4
ระดับนักเรียน													
เพศ	✓			✓	✓	✓		✓	✓			✓	7
อายุ	✓			✓				✓				✓	4
การย้ายถิ่น/แหล่งอาศัย								✓					1
เชื้อชาติ/ชาติกำเนิด							✓	✓	✓				3
เศรษฐานะของครอบครัว	✓	✓		✓	✓	✓	✓				✓	✓	8
คุณลักษณะของครอบครัว	✓				✓	✓							3
ความรู้เดิมของนักเรียน/ผลสัมฤทธิ์การเรียนที่ผ่านมา		✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	10
ความเชื่อมั่นในตนเอง/เชิงวิชาการ	✓			✓		✓						✓	4
เจตคติและแรงจูงใจ	✓			✓		✓						✓	4

ผลการสังเคราะห์กลุ่มตัวแปรซึ่งใช้เป็นตัวบ่งชี้คุณภาพการจัดการศึกษาของสถานศึกษา มีความสอดคล้องกับตัวแปรต่างๆ ใน PISA 2009 และเมื่อนำมาพิจารณาตัวแปรที่ศึกษามากที่สุด ในระดับสถานศึกษาได้แก่ ขนาดสถานศึกษา และในระดับตัวแปรนักเรียนได้แก่ ความรู้เดิมของนักเรียน/ผลสัมฤทธิ์ทางการเรียนที่ผ่านมา และแหล่งข้อมูลของโครงการประเมินผลนักเรียนนานาชาติ PISA 2009 ที่นำมาใช้ในการวิจัยครั้งนี้ สรุปรายละเอียดตามตารางที่ 15

ตารางที่ 15 อันดับปัจจัยที่มีผลต่อคุณภาพการจัดการศึกษาที่สอดคล้องกับโครงการประเมินผลนักเรียนนานาชาติ 2009

อันดับ	ตัวแปร	แหล่งข้อมูลใน PISA 2009
ระดับสถานศึกษา		
1	ขนาดสถานศึกษา	ขนาดสถานศึกษา ขนาดห้องเรียน
2	ที่ตั้งสถานศึกษา	ประเภทของสถานศึกษา ตำแหน่งที่ตั้งของสถานศึกษา
3	งบประมาณ/รายได้ของหน่วยงาน	งบประมาณที่ได้รับ
	การขาดแคลนครู	ระดับการขาดแคลนครูที่มีคุณวุฒิ สัดส่วนนักเรียนต่อครู
4	แหล่งทรัพยากรอุปกรณ์	ทรัพยากรการเรียน ความรับผิดชอบเกี่ยวกับการจัดสรรทรัพยากร
	หลักสูตรและการประเมินผล	หลักสูตรและการประเมินผล
5	ประสบการณ์สอนของครู	การจัดกิจกรรมเพื่อการเรียนรู้
6	บริบทสถานศึกษา/สังกัด	สังกัดของสถานศึกษา
ระดับนักเรียน		
1	ความรู้เดิมของนักเรียน/ผลสัมฤทธิ์การเรียนที่ผ่านมา	ผลสัมฤทธิ์ทางการเรียนที่ผ่านมา
2	เศรษฐานะของครอบครัว	เศรษฐานะของครอบครัว แหล่งทรัพยากรการศึกษาที่บ้าน ความมั่งคั่งของครอบครัว
3	เพศ	เพศ
4	ความเชื่อมั่นในตนเอง/เชิงวิชาการ	การใช้เวลาเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา การใช้เวลาในการศึกษาวิทยาศาสตร์ด้วยตนเอง

ตารางที่ 15 อันดับปัจจัยที่มีผลต่อคุณภาพการจัดการศึกษาที่สอดคล้องกับโครงการประเมินผลนักเรียนนานาชาติ ปี 2009 (ต่อ)

อันดับ	ตัวแปร	แหล่งข้อมูลใน PISA 2009
ระดับนักเรียน (ต่อ)		
5	เจตคติและแรงจูงใจ	ความสนใจในการเรียนรู้วิทยาศาสตร์ ความเชื่อมั่นในประสิทธิภาพของตนเองในการเรียนวิทยาศาสตร์ ความเชื่อในความสามารถของตนเองในการเรียนวิทยาศาสตร์ ความเพลิดเพลินทางวิทยาศาสตร์ แรงจูงใจภายนอกในการเรียนวิทยาศาสตร์ การเห็นคุณค่าโดยทั่วไปในวิทยาศาสตร์

นอกจากตัวแปรที่ใช้ในการศึกษาแล้ว ผู้วิจัยได้สังเคราะห์ตัวแปรที่ใช้ในการศึกษาการทำหน้าที่ต่างกันข้อสอบ (DIF) และการทำหน้าที่ต่างกันของแบบสอบ (DTF) จากผลการวิจัยของรัชนก ยี่สุ่นศรี (2544), Tuksino, Kanjanawasee, และ Pasiphol (2010), Sirikit, Pasiphol, และ Kanjanawasee (2012), Le (2009) Liu และ Wilson (2009), Yildirim และYildirim (2011) Fleischman และคณะ (2010) ที่สอดคล้องกับ PISA 2009 พบว่า ตัวแปรที่มีการนำมาศึกษาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ ได้แก่ เพศ การเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา เศรษฐานะของครอบครัว แหล่งทรัพยากรการศึกษาที่บ้าน และความมั่งคั่งของครอบครัว ดังนั้นในการวิจัยครั้งนี้ผู้วิจัยจึงนำตัวแปรดังกล่าวเข้ามาศึกษาการทำหน้าที่ต่างกันของทำหน้าที่ต่างกันของข้อสอบ (DIF) และแบบสอบ (DTF) และการทำหน้าที่ต่างกันเป็นขั้น (DSF)

ตอนที่ 5 กรอบแนวคิดในการวิจัย

จากการศึกษาค้นคว้าเอกสารและงานวิจัยที่เกี่ยวข้องทั้งในประเทศและต่างประเทศ พบว่าข้อสอบ/แบบสอบที่ดีต้องมีความยุติธรรมกับผู้เข้าสอบ ไม่โน้มเอียงเข้าหากลุ่มใดกลุ่มหนึ่งเมื่อผู้สอบมีความสามารถอยู่ในระดับเท่ากัน นอกจากนี้การประเมินผลคะแนนการสอบจะต้องคำนึงถึงความแตกต่างของปัจจัยที่เกี่ยวข้องในระดับต่างๆ เช่น ตัวแปรระดับนักเรียนหรือระดับสถานศึกษาร่วมด้วย จากเหตุผลดังกล่าว ผู้วิจัยจึงกำหนดกรอบแนวคิดในการวิจัยเพื่อหาคะแนนมูลค่าเพิ่มของสถานศึกษา โดยพิจารณาจากผลการสังเคราะห์ตัวแปรที่ใช้ในการศึกษาแบ่งเป็น 2 ระดับคือ ตัวแปรประชากรระดับสถานศึกษาและตัวแปรประชากรระดับนักเรียน โดยที่

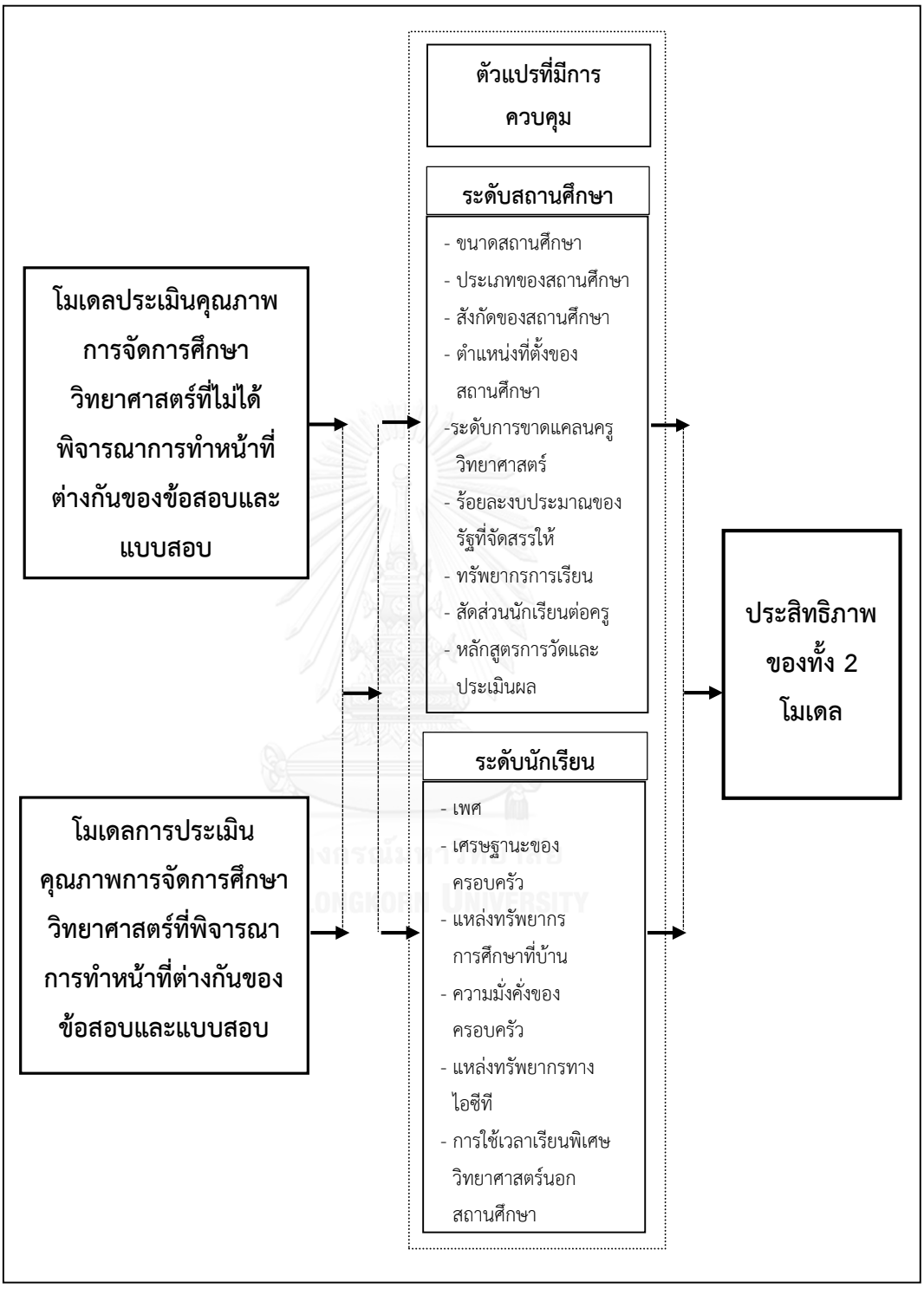
ตัวแปรประชากรระดับสถานศึกษา ได้แก่ ขนาดสถานศึกษา ประเภทของสถานศึกษา สังกัดของสถานศึกษา ตำแหน่งที่ตั้งของสถานศึกษา ระดับการขาดแคลนครูวิทยาศาสตร์ ร้อยละงบประมาณของรัฐที่จัดสรรให้ ทรัพยากรการเรียน สัดส่วนนักเรียนต่อครู หลักสูตรการวัดและประเมินผล

ตัวแปรประชากรระดับนักเรียน ได้แก่ เพศ เศรษฐานะของครอบครัว แหล่งทรัพยากรการศึกษาที่บ้าน ความมั่งคั่งของครอบครัว แหล่งทรัพยากรทางไอซีที การใช้เวลาเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา

และตัวแปรที่ใช้ศึกษาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ และการทำหน้าที่ต่างกันเป็นขั้น ได้แก่ เพศ การเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา เศรษฐานะของครอบครัว แหล่งทรัพยากรการศึกษาที่บ้าน ความมั่งคั่งของครอบครัว ส่วนโมเดลมูลค่าเพิ่มพหุระดับประกอบไปด้วย โมเดลที่ 1 เป็นโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ที่ไม่ได้พิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ ส่วนโมเดลการวิเคราะห์มูลค่าเพิ่มพหุระดับเป็นโมเดลที่มีการควบคุมแปรระดับนักเรียนและสถานศึกษา และโมเดลที่ 2 เป็นโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ที่พิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ โดยการตัดข้อสอบที่มีการทำหน้าที่ต่างกันของข้อสอบและการทำหน้าที่ต่างกันเป็นขั้น ส่วนโมเดลการวิเคราะห์มูลค่าเพิ่มพหุระดับเป็นโมเดลที่โมเดลที่มีการควบคุมตัวแปรระดับนักเรียนและสถานศึกษา รายละเอียดกรอบแนวคิดในการวิจัยแสดงรายละเอียดได้ดังนี้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

กรอบแนวคิดในการวิจัย



บทที่ 3
วิธีดำเนินการวิจัย

วิธีดำเนินการวิจัยครั้งนี้ ผู้วิจัยแบ่งเนื้อหาการนำเสนอเป็น 5 ตอน ได้แก่ ตอนที่หนึ่ง คือ ประชากรและกลุ่มตัวอย่าง ตอนที่สอง คือ รูปแบบของแบบสอบประเมินการรู้เรื่องด้านวิทยาศาสตร์และรูปแบบข้อสอบ ตอนที่สาม คือ คุณภาพเครื่องมือที่ใช้ในการวิจัย ตอนที่สี่ คือ ปัจจัยหรือตัวแปรที่นำมาใช้ในการวิจัย และตอนสุดท้าย คือ การวิเคราะห์ข้อมูล รายละเอียดมีดังนี้

ตอนที่ 1 ประชากรและกลุ่มตัวอย่าง

การวิจัยครั้งนี้เป็นการศึกษาจากข้อมูลทุติยภูมิของโครงการประเมินผลนักเรียนนานาชาติ ปี 2552 หรือ PISA 2009 จากประชากรนักเรียนของประเทศไทยที่มีช่วงอายุ 15 ปี 2 เดือน จนถึง 16 ปี 3 เดือน กลุ่มตัวอย่างของการประเมินถูกเลือกการสุ่มแบบหลายขั้น (Multi stage random sampling) ครอบคลุมทุกพื้นที่ภูมิศาสตร์ของประเทศ 7 พื้นที่ และสังกัดโรงเรียน 8 กลุ่ม โดยพิจารณาให้มีปริมาณของพื้นที่และจำนวนโรงเรียนใกล้เคียงกันมากที่สุด รายละเอียดพื้นฐานกลุ่มตัวอย่างตามกรอบการสุ่มกลุ่มตัวอย่างจำแนกตามภาคพื้นที่ภูมิศาสตร์และตามสังกัดโรงเรียน รายละเอียดตามตารางที่ 16

ตารางที่ 16 กรอบการสุ่มกลุ่มตัวอย่างจำแนกตามภาคพื้นที่ภูมิศาสตร์และสังกัดโรงเรียน

กรอบการสุ่ม	จำแนกกลุ่มการสุ่ม	
	ภาคพื้นที่ภูมิศาสตร์	สังกัดโรงเรียน
รายละเอียดพื้นฐานกลุ่มตัวอย่าง	1. กรุงเทพมหานครและปริมณฑล 2. ภาคกลาง 3. ภาคเหนือตอนบน 4. ภาคเหนือตอนล่าง 5. ภาคอีสานตอนบน 6. ภาคอีสานตอนล่าง 7. ภาคใต้	1. สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน สำหรับโรงเรียนซึ่งเคยเป็นโรงเรียนขยายโอกาส (สพฐ.1) 2. สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน สำหรับโรงเรียนในสังกัดกรมสามัญศึกษาเดิม (สพฐ.2) 3. สังกัดสำนักบริหารงานคณะกรรมการส่งเสริมการศึกษาเอกชน(สช.) 4. สังกัดสำนักงานการศึกษากรุงเทพมหานคร (กทม.) 5. สังกัดสำนักประสานและพัฒนาการจัดการศึกษาท้องถิ่น (กศท.) 6. โรงเรียนสาธิตของมหาวิทยาลัย (สาธิต) 7. โรงเรียนอาชีวศึกษาเอกชน (อศ.1) 8. สถาบันอาชีวศึกษาของรัฐ (อศ.2)

เมื่อได้จำนวนพื้นที่ภูมิศาสตร์และสังกัดโรงเรียน จึงดำเนินการสุ่มจำนวนโรงเรียน ตามสังกัดโรงเรียนและภาคพื้นที่ภูมิศาสตร์ จากจำนวนโรงเรียนทั้งหมด 11,863 โรงเรียน สุ่มมาได้ 230 โรงเรียน โดยจำแนกตามพื้นที่ภูมิศาสตร์และสังกัดโรงเรียน รายละเอียดตามตารางที่ 17

ตารางที่ 17 จำนวนโรงเรียนกลุ่มตัวอย่าง PISA 2009(สังกัดโรงเรียน x พื้นที่ภูมิศาสตร์)

พื้นที่ภูมิศาสตร์	ประเภทโรงเรียน								รวม
	สพฐ.1	สพฐ.2	สช.	กทม.	กศท.	สาธิต	อศ.1	อศ.2	
กทม.และปริมณฑล	1	15	3	22	1	7	5	1	**55**
ภาคกลาง	3	5	1	-	2	1	1	2	**15**
ภาคเหนือตอนบน	4	8	1	-	-	1	1	2	**17**
ภาคเหนือตอนล่าง	5	8	-	-	2	1	-	2	**18**
ภาคอีสานตอนบน	8	18	-	-	2	3	2	3	**36**
ภาคอีสานตอนล่าง	9	15	1	-	7	-	1	3	**36**
ภาคใต้	4	11	6	-	2	1	2	3	**29**
ภาคตะวันออก	2	4	-	-	-	1	1	1	**9**
ภาคตะวันตก	2	7	1	-	2	1	-	2	**15**
รวม	**38**	**91**	**13**	**22**	**18**	**16**	**13**	**19**	**230**

ดำเนินการสุ่มนักเรียนกลุ่มตัวอย่างจำแนกตามสังกัดสถานศึกษา ได้จำนวนนักเรียนที่สุ่มได้ทั้งหมด 6,612 คน โดยมีจำนวนนักเรียนที่เข้าสอบจริง 6,225 คน รายละเอียดจำนวนกลุ่มตัวอย่างนักเรียน PISA 2009 จำแนกตามสังกัดสถานศึกษารายละเอียดตามตารางที่ 18

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 18 จำนวนกลุ่มตัวอย่างนักเรียนในโครงการประเมินผลนานาชาติจำแนกตามสังกัด

สถานศึกษาในสังกัด	สถานศึกษากลุ่มตัวอย่าง	นักเรียนกลุ่มตัวอย่าง	นักเรียนที่เข้าสอบจริง
สพฐ1. (ขยายโอกาสเดิม)	38	397	384
สพฐ2. (สามัญฯ เดิม)	91	3,177	3,014
สช.(เอกชน)	13	425	399
กทม.(กรุงเทพมหานคร)	22	508	478
กศท.(การศึกษาท้องถิ่น)	18	514	483
สาธิต (ของมหาวิทยาลัย)	16	501	477
อศ.1(อาชีวของเอกชน)	13	442	405
อศ.2(อาชีวของรัฐบาล)	19	648	585
รวม	**230**	**6,612**	**6,225**

ตอนที่ 2 รูปแบบของแบบสอบประเมินการรู้เรื่องด้านวิทยาศาสตร์และรูปแบบข้อสอบ

กรอบการประเมินวิทยาศาสตร์ครอบคลุมสมรรถนะทางวิทยาศาสตร์ 3 สมรรถนะ ได้แก่ การอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ การระบุประเด็นทางวิทยาศาสตร์ และการใช้ประจักษ์พยานทางวิทยาศาสตร์ และครอบคลุมความรู้ทางวิทยาศาสตร์ 7 ด้าน ได้แก่ ระบบโลกและอวกาศ ระบบการดำรงชีวิต ระบบทางกายภาพ ระบบเทคโนโลยี วิทยาศาสตร์และเทคโนโลยีในสังคม การอธิบายเชิงวิทยาศาสตร์ และการแสวงหาความรู้เชิงวิทยาศาสตร์ ทั้งนี้ PISA 2009 มีการทดสอบกับนักเรียนแตกต่างกันหลายกลุ่ม จึงมีการสร้างข้อสอบให้มีความหลายหลายและเลือกให้แต่ละแบบมีจำนวนข้อที่ใกล้เคียงกัน ได้แก่ แบบเลือกตอบทั้งเลือกตอบธรรมดา และเลือกตอบเชิงซ้อน ตอบแบบอิสระโดยสร้างคำตอบแบบเปิด และสร้างคำตอบแบบปิด

โดยที่ข้อสอบส่วนใหญ่ที่มีลักษณะข้อสอบแบบเลือกตอบธรรมดา ข้อสอบแบบเลือกตอบเชิงซ้อน และข้อสอบแบบตอบแบบอิสระโดยสร้างคำตอบแบบปิดมีการตรวจให้คะแนนแบบทวิวิภาค ส่วนตอบแบบอิสระโดยสร้างคำตอบแบบเปิดมีการตรวจให้คะแนนทั้งแบบทวิวิภาคและพหุวิภาค รายละเอียดจำนวนข้อคำถามจำแนกตามประเภทข้อสอบรายละเอียดตามตารางที่ 19

ตารางที่ 19 ลักษณะข้อสอบที่ใช้ในจำแนกตามสมรรถนะและความรู้ทางวิทยาศาสตร์

สมรรถนะทางคณิตศาสตร์	เลือกตอบ (ข้อ)	เลือกตอบเชิงซ้อน(ข้อ)	ตอบแบบอิสระแบบเปิด (ข้อ)	ตอบแบบอิสระแบบปิด (ข้อ)	ข้อสอบรวม(ข้อ)
สมรรถนะทางวิทยาศาสตร์					
การอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์	9	7	7	1	**24**
การระบุประเด็นทางวิทยาศาสตร์	3	6	3	-	**12**
การใช้ประจักษ์พยานทางวิทยาศาสตร์	6	4	7	-	**17**
รวม	**18**	**17**	**17**	**1**	**53**

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 19 ลักษณะข้อสอบที่ใช้ในจำแนกตามสมรรถนะและความรู้ทางวิทยาศาสตร์ (ต่อ)

สมรรถนะทางคณิตศาสตร์	เลือกตอบ (ข้อ)	เลือกตอบเชิงซ้อน(ข้อ)	ตอบแบบอิสระแบบเปิด (ข้อ)	ตอบแบบอิสระแบบปิด (ข้อ)	ข้อสอบรวม(ข้อ)
ความรู้ทางวิทยาศาสตร์					
ความรู้วิทยาศาสตร์					
ระบบโลกและอวกาศ	3	2	3	-	**8**
ระบบการดำรงชีวิต	2	3	4	-	**9**
ระบบทางกายภาพ	4	2	-	1	**7**
ระบบเทคโนโลยี	1	2	1	-	**4**
วิทยาศาสตร์และเทคโนโลยีในสังคม	-	1	1	-	**2**
ความรู้เกี่ยวกับวิทยาศาสตร์					
การอธิบายเชิงวิทยาศาสตร์	5	2	5	-	**12**
การแสวงหาความรู้เชิงวิทยาศาสตร์	3	5	3	-	**11**
รวม	**18**	**17**	**17**	**1**	**53**

ตอนที่ 3 คุณภาพเครื่องมือที่ใช้ในการวิจัย

คุณภาพเครื่องมือที่ใช้ในการประเมินผล PISA 2009 จากรายงานของ OECD (2012) มีการรายงานผลความเที่ยงของเครื่องมือที่ใช้ในการทดสอบการรู้เรื่องวิทยาศาสตร์ คณิตศาสตร์ และการอ่าน พบว่า ความเที่ยงของเครื่องมือจากทุกประเทศมีค่าอยู่ระหว่าง .882 - .921 แบบทดสอบการรู้เรื่องวิทยาศาสตร์มีค่าความเที่ยง เท่ากับ .896 สำหรับประเทศไทยความเที่ยงของเครื่องมือมีค่าอยู่ระหว่าง .84 - .89 และแบบทดสอบการรู้เรื่องวิทยาศาสตร์มีค่าความเที่ยงเท่ากับ .84

ทั้งนี้ OECD (2012) ได้ศึกษาองค์ประกอบที่ประกอบด้วยสามองค์ประกอบหลักจากดัชนีทางเศรษฐกิจ สังคม และสถานทางวัฒนธรรม หรือเศรษฐานะของครอบครัว (ESCS) ประกอบด้วยอาชีพของผู้ปกครองสูงสุด (HISEI) การศึกษาสูงสุดของผู้ปกครอง (PARED) ดัชนีความเป็นเจ้าของที่อยู่อาศัย (HOMEPOS) จากการศึกษาความคงที่ข้ามประเทศ (consistency across countries) โดยการวิเคราะห์องค์ประกอบหลัก (PCA) เพื่อหาน้ำหนักองค์ประกอบของประเทศที่เข้าร่วมโดยพิจารณาความสัมพันธ์ตัวแปรทั้งสาม พบว่า ประเทศไทยมีน้ำหนักองค์ประกอบอาชีพของผู้ปกครองสูงสุด (HISEI) เท่ากับ .86 การศึกษาสูงสุดของผู้ปกครอง (PARED) เท่ากับ .87 ดัชนีความเป็น

เจ้าของที่อยู่อาศัย (HOMEPOS) เท่ากับ .84 ตามลำดับ และค่าความเที่ยงของ Cronbach's alpha เท่ากับ .81 ซึ่งใกล้เคียงกับน้ำหนักองค์ที่เป็นค่ามัธยฐานจากกลุ่มประเทศที่เป็นสมาชิก OECD และประเทศที่เข้าร่วมประเมิน รายละเอียดตามตารางที่ 20

ตารางที่ 20 น้ำหนักองค์ประกอบและความเที่ยง

น้ำหนักองค์ประกอบและความสอดคล้องภายใน	น้ำหนักขององค์ประกอบย่อย			ความเที่ยง
	HISEI	PARED	HOMEPOS	
ประเทศไทย	.86	.87	.84	.81
มัยฐานของของกลุ่ม OECD	.80	.79	.70	.65
มัยฐานของประเทศที่เข้าร่วม	.81	.80	.75	.69

นอกจากนี้น้ำหนักองค์ประกอบย่อยสามประกอบทุกรอบการประเมิน พบว่า องค์ประกอบย่อยทั้งสามมีน้ำหนักองค์ประกอบที่ใกล้เคียงกันในทุกๆ รอบที่มีการประเมิน ซึ่งในปี 2009 น้ำหนักองค์ประกอบอาชีพของผู้ปกครองสูงสุด (HISEI) เท่ากับ 0.81 การศึกษาสูงสุดของผู้ปกครอง (PARED) เท่ากับ 0.81 และดัชนีความเป็นเจ้าของที่อยู่อาศัย (HOMEPOS) เท่ากับ 0.74 ตามลำดับ (OECD, 2012) รายละเอียดตามตารางที่ 21

ตารางที่ 21 องค์ประกอบของ ESCS ในปี 2000, 2003, 2006, 2009

รอบการประเมินของ PISA	น้ำหนักขององค์ประกอบย่อย		
	HISEI	PARED	HOMEPOS
2000	0.81	0.79	0.75
2003	0.81	0.81	0.77
2006	0.81	0.81	0.75
2009	0.81	0.81	0.74

ตอนที่ 4 ปัจจัยหรือตัวแปรที่นำมาใช้ในการวิจัย

โครงสร้างข้อมูลโครงการประเมินผลนักเรียนนานาชาติ ปี 2009 มีลักษณะโครงสร้างแบบลดหลั่น คือ นักเรียนแทรกอยู่ในสถานศึกษา ดังนั้นผู้วิจัยจึงแบ่งตัวแปรออกเป็นระดับสถานศึกษาและระดับนักเรียน โดยพิจารณาจากข้อคำถามที่ใช้ในการเก็บจากกลุ่มตัวอย่างกับผู้บริหารและนักเรียน ซึ่งจำแนกตัวแปรที่ใช้ในการศึกษาเป็น 2 มิติ คือ ตัวแปรที่ศึกษาพหุระดับ และตัวแปรที่นำมาศึกษาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ และการทำหน้าที่ต่างกันเป็นขั้น ทั้งนี้ตัว

แปรที่มีการศึกษาพหุระดับจำแนกได้เป็น ตัวแปรประชากรระดับสถานศึกษาและตัวแปรประชากรระดับนักเรียน ดังนี้

ตัวแปรประชากรระดับสถานศึกษา ได้แก่ ขนาดสถานศึกษา ประเภทของสถานศึกษา สังกัดของสถานศึกษา ตำแหน่งที่ตั้งของสถานศึกษา ระดับการขาดแคลนครูวิทยาศาสตร์ ร้อยละงบประมาณของรัฐที่จัดสรรให้ ทรัพยากรการเรียน สัดส่วนนักเรียนต่อครู และหลักสูตรการวัดและประเมินผล

ตัวแปรประชากรระดับนักเรียน ได้แก่ เพศ เศรษฐานะของครอบครัว แหล่งทรัพยากรการศึกษาที่บ้าน ความมั่งคั่งของครอบครัว แหล่งทรัพยากรทางไอซีที และการใช้เวลาเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา รายละเอียดสเกลการวัดและความหมายตัวแปรตามตารางที่ 22

ตารางที่ 22 สเกลการวัดและความหมายตัวแปรที่ใช้ในการศึกษา

ตัวแปร	สเกลการวัดและความหมาย
ตัวแปรระดับสถานศึกษา	
1. ขนาดสถานศึกษา	ตัวแปรจัดประเภท แสดงจำนวนนักเรียนในสถานศึกษา จำแนกเป็นขนาดใหญ่ (นักเรียนมากกว่า 1,500 คน), ขนาดปานกลาง (นักเรียน 500 – 1,500 คน), และขนาดเล็ก (นักเรียน 1– 499 คน)
2. ประเภทของถานศึกษา	ตัวแปรจัดประเภท จำแนกเป็น สถานศึกษาของรัฐและเอกชน
3. สังกัดของสถานศึกษา	ตัวแปรจัดประเภท (เป็นตัวแปรที่เก็บเพิ่มจากหน่วยงาน สสวท.) จำแนกเป็นสถานศึกษา 8 สังกัด ได้แก่ สพฐ.1, สพฐ.2, สช., สกอ., กศท., กทม., อศ.1, และ อศ.2
4. ตำแหน่งที่ตั้งของสถานศึกษา	ตัวแปรจัดประเภท แสดงขนาดของชุมชนที่สถานศึกษาตั้งอยู่ จำแนกเป็น ตำแหน่งหมู่บ้าน (ประชาชนน้อยกว่า 3,000 คน), เมืองเล็ก (ประชาชน 3,000-15,000 คน), เมือง (ประชาชน 15,000 -100,000 คน), เมืองใหญ่ (ประชาชน 100,000 -1,000,000 คน), และเมืองใหญ่ (ประชาชนมากกว่า 1,000,000 คน)
5. ระดับการขาดแคลนครูวิทยาศาสตร์	ตัวแปรจัดประเภท แสดงระดับการขาดแคลนครูวิทยาศาสตร์ จำแนกเป็น 4 ระดับคือ ไม่ขาดแคลน, ขาดแคลนน้อย, ขาดแคลนเป็นบางส่วน, และขาดแคลนจำนวนมาก
6. ร้อยละงบประมาณของรัฐที่จัดสรรให้สถานศึกษา	ตัวแปรต่อเนื่อง คิดจากร้อยละงบประมาณของรัฐที่จัดสรรให้กับสถานศึกษา

ตารางที่ 22 สเกลการวัดและความหมายตัวแปรที่ใช้ในการศึกษา (ต่อ)

ตัวแปร	สเกลการวัดและความหมาย
7. ทรัพยากรการเรียน	ตัวแปรต่อเนื่อง เป็นค่าดัชนีชี้วัดที่ได้มาจากข้อรายการพื้นฐาน 7 ประการที่ใช้วัดองค์ประกอบตามศักยภาพ ด้านอุปกรณ์การเรียนการสอนของโรงเรียน ได้แก่ ห้องแล็บวิทยาศาสตร์ คู่มือการเรียนการสอน คอมพิวเตอร์ อินเตอร์เน็ต ซอฟแวร์ที่ใช้ในการเรียนการสอน วัสดุอุปกรณ์ในห้องสมุด และโสตทัศนวัสดุ
8. สัดส่วนนักเรียนต่อครู	ตัวแปรต่อเนื่อง เป็นค่าที่แสดงสัดส่วนนักเรียนต่อครูของสถานศึกษา
9. หลักสูตรการวัดและประเมินผล	ตัวแปรต่อเนื่อง เป็นค่าดัชนีที่ได้มาจากคะแนนรวมรายงานผู้ที่มีความรับผิดชอบด้านหลักสูตรและการวัดและประเมินผลของโรงเรียน ได้แก่ นโยบายการประเมินผลนักเรียน คู่มือที่ใช้เนื้อหาในหลักสูตร และหลักสูตรที่สนับสนุน
ตัวแปรระดับนักเรียน	
1. คะแนนเสกลของการประเมินผลการรู้เรื่องวิทยาศาสตร์	ตัวแปรต่อเนื่อง เป็นคะแนนที่ได้จากการแปลงค่าประมาณความสามารถของนักเรียนเป็นคะแนนสเกลของ PISA ในวิชาวิทยาศาสตร์
2. เพศ	ตัวแปรจัดประเภท จำแนกเป็น นักเรียนชายหรือนักเรียนหญิง
3. เศรษฐานะของครอบครัว	ตัวแปรต่อเนื่อง เป็นดัชนีชี้วัดที่ได้มาจากจากคะแนนรวมที่วัดได้จากสถานะทางการงาน การศึกษาของพ่อแม่ ฐานะทางเศรษฐกิจ ทรัพยากรด้านการศึกษาและวัฒนธรรมของครอบครัวของเด็กนักเรียน
4. แหล่งทรัพยากรการศึกษาที่บ้าน	ตัวแปรต่อเนื่อง ดัชนีชี้วัดที่ได้มาจากคะแนนรวมที่วัดได้จากรายการสิ่งของที่มีอยู่ภายในบ้าน เช่น พจนานุกรม คอมพิวเตอร์ โต๊ะทำงาน หนังสือเรียน
5. ความมั่งคั่งของครอบครัว	ตัวแปรต่อเนื่อง เป็นค่าดัชนีชี้วัดที่ได้มาจากคะแนนรวมที่วัดได้จากสิ่งของเครื่องใช้ต่างๆ ภายในบ้าน เช่น อินเทอร์เน็ต เครื่องล้างจาน เครื่องเล่นดีวีดี โทรศัพท์เคลื่อนที่ โทรทัศน์ รถ และห้องน้ำในบ้านที่มีฝักบัวหรืออ่างอาบน้ำ
6. แหล่งทรัพยากรทางไอซีที	ตัวแปรต่อเนื่อง เป็นค่าดัชนีชี้วัดที่ได้มาจากคะแนนรวมแหล่งทรัพยากรทางไอซีทีของนักเรียน
7. การใช้เวลาเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา	ตัวแปรจัดประเภท แสดงระดับของการใช้เวลาเรียนในการเรียนพิเศษของนักเรียนในหนึ่งสัปดาห์ แบ่งเป็น 5 ระดับ คือ ไม่มีเวลา, น้อยกว่า 2 ชม., 2-4 ชม., 4-6 ชม., และมากกว่า 6 ชม.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

โดยที่ตัวแปรที่นำมาศึกษาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ และการทำหน้าที่ต่างกันเป็นขั้น จำนวน 5 ตัวแปร ได้แก่ เพศ เศรษฐานะของครอบครัว แหล่งทรัพยากรการศึกษาที่บ้าน ความมั่งคั่งของครอบครัว และการเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา รายละเอียดสัญลักษณ์และความหมายของตัวแปรที่นำมาศึกษาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ และการทำหน้าที่ต่างกันเป็นขั้นตามตารางที่ 23

ตารางที่ 23 สัญลักษณ์และความหมายตัวแปรที่ใช้ในการศึกษาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ

ตัวแปร	สัญลักษณ์	ความหมาย
เพศ	GENDER	เพศของนักเรียน ซึ่งนักเรียนหญิงเป็นกลุ่มอ้างอิง (reference group) = 1 และนักเรียนชายเป็นกลุ่มเปรียบเทียบ (focal group) = 2
เศรษฐานะของครอบครัว	ESCS	เศรษฐานะของครอบครัว ซึ่งนักเรียนที่มีเศรษฐานะของครอบครัวสูงกว่าค่าเฉลี่ยเป็นกลุ่มอ้างอิง (reference group) = 1 และนักเรียนที่มีเศรษฐานะของครอบครัวต่ำกว่าค่าเฉลี่ยเป็นกลุ่มเปรียบเทียบ (focal group) = 2
แหล่งทรัพยากรการศึกษาที่บ้าน	HEDRES	แหล่งทรัพยากรการศึกษาที่บ้าน ซึ่งนักเรียนที่มีแหล่งทรัพยากรการศึกษาที่บ้านสูงกว่าค่าเฉลี่ยเป็นกลุ่มอ้างอิง (reference group) = 1 และนักเรียนที่มีแหล่งทรัพยากรการศึกษาที่บ้านต่ำกว่าค่าเฉลี่ยเป็นกลุ่มเปรียบเทียบ (focal group) = 2
ความมั่งคั่งของครอบครัว	WEALTH	ความมั่งคั่งของครอบครัว ซึ่งนักเรียนที่มีความมั่งคั่งของครอบครัวสูงกว่าค่าเฉลี่ยเป็นกลุ่มอ้างอิง (reference group) = 1 และนักเรียนที่มีความมั่งคั่งของครอบครัวต่ำกว่าค่าเฉลี่ยเป็นกลุ่มเปรียบเทียบ (focal group) = 2
การใช้เวลาเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา	EXTRA	การใช้เวลาเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา ซึ่งนักเรียนที่เรียนพิเศษเป็นกลุ่มอ้างอิง (reference group) = 1 และนักเรียนที่ไม่เรียนพิเศษเป็นกลุ่มเปรียบเทียบ (focal group) = 2

ตอนที่ 5 การวิเคราะห์ข้อมูล

ส่วนการวิเคราะห์ข้อมูลผู้วิจัยแบ่งการวิเคราะห์ออกเป็น 6 ขั้นตอน ประกอบด้วย ขั้นตอนที่หนึ่ง คือ ตรวจสอบข้อสอบและแบบสอบที่นำมาศึกษา ขั้นตอนที่สอง คือ วิเคราะห์ข้อมูลเบื้องต้นตามตัวแปรที่ศึกษา ขั้นตอนที่สาม คือ ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและแบบสอบและ การทำหน้าที่ต่างกันเป็นขั้น ขั้นตอนที่สี่ คือ ตรวจสอบเงื่อนไขการวิเคราะห์พหุระดับและ

ความเป็นเอกมิติของแบบสอบและการวิเคราะห์พหุระดับโดยใช้โมเดลมูลค่าเพิ่มพหุระดับ ขั้นตอนที่ห้า คือ พิจารณาการตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบและโครงสร้างข้อสอบทั้งก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน และตอนสุดท้าย คือ เปรียบเทียบประสิทธิภาพโมเดลการประเมินคุณภาพการจัดการศึกษาวิชาวิทยาศาสตร์ รายละเอียดดังนี้

5.1 ตรวจสอบข้อสอบและแบบสอบที่นำมาศึกษา ในส่วนขั้นตอนการตรวจสอบข้อสอบและแบบสอบที่นำมาศึกษาได้มาจากฐานข้อมูล (data base) ของโครงการประเมินผลนักเรียนนานาชาติ ปี 2009 เนื้อหาสาระการรู้เรื่องวิทยาศาสตร์ มีแบบสอบที่ใช้จริงจำนวน 9 ฉบับจากแบบสอบ 13 ฉบับ ได้แก่ แบบสอบฉบับที่ 2, 3, 4, 7, 8, 9, 10, 12, และ 13 ต่อมาผู้วิจัยได้พิจารณาลักษณะการตอบของนักเรียนในแบบสอบแต่ละฉบับโดยคัดเลือกนักเรียนที่มีลักษณะการตอบครบถ้วนและไม่เกิดการสูญหายของข้อมูล (missing data) เพื่อความเหมาะสมต่อการวิเคราะห์ข้อมูลและลดความคลาดเคลื่อนอันเกิดจากผลการตอบของนักเรียน เหลือจำนวนนักเรียนที่ใช้ในการวิจัยครั้งนี้ทั้งหมด 4,292 คน จากแบบสอบ 9 ฉบับ รายละเอียดตามตารางที่ 24

ตารางที่ 24 จำนวนนักเรียนก่อนการคัดเลือกและที่ใช้ในการวิจัย

ฉบับที่	จำนวนนักเรียนถูกสุ่มในแบบสอบ	จำนวนนักเรียนที่ใช้จริง
2	486	484
3	475	475
4	478	473
7	480	476
8	483	475
9	473	473
10	471	471
12	478	478
13	487	487
รวม	**4,311**	**4,292**

5.2 วิเคราะห์ข้อมูลเบื้องต้นตามตัวแปรที่ศึกษา ในส่วนนี้ผู้วิจัยดำเนินการวิเคราะห์ค่าสถิติพื้นฐานของข้อมูลด้วยโปรแกรมวิเคราะห์ SPSS 16.0 for windows โดยเป็นการวิเคราะห์ข้อมูลเบื้องต้นด้วยสถิติภาคบรรยาย ได้แก่ ความถี่ ร้อยละ ค่าเฉลี่ย ค่าต่ำสุด ค่าสูงสุด ช่วงห่างระหว่างค่าสูงสุดและต่ำสุด ส่วนเบี่ยงเบนมาตรฐาน ความเบ้และความโด่ง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

5.3 ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและแบบสอบและการทำหน้าที่ต่างกันเป็นขั้น ในส่วนนี้เป็นการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและแบบสอบและการทำหน้าที่ต่างกันเป็นขั้น เพื่อนำผลที่ได้มาใช้ในการพิจารณาคัดเลือกข้อสอบที่มีความเหมาะสม ทั้งนี้ในการตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน ผู้วิจัยได้อาศัยเกณฑ์ที่ได้กำหนดขึ้นมาเพื่อการพิจารณาตัดข้อสอบ ต่อจากนั้นจึงทดสอบการทำหน้าที่ต่างกันทั้งฉบับหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันและตรวจสอบโครงสร้างแบบสอบด้วยการตรวจสอบความตรงเชิงโครงสร้างด้วยการวิเคราะห์องค์ประกอบเชิงยืนยันแล้วจึงนำคะแนนผลการสอบที่ได้ไปหาค่าความสามารถของผู้สอบอีกครั้ง ดังนั้นขั้นตอนนี้ผู้วิจัยจึงแบ่งผลการวิเคราะห์เป็น 3 ส่วนย่อย ได้แก่ ส่วนที่หนึ่ง คือ ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและแบบสอบและการทำหน้าที่ต่างกันเป็นขั้นตามตัวแปรที่ศึกษา ส่วนที่สอง คือ พิจารณาการตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันและโครงสร้างข้อสอบทั้งก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน และส่วนสุดท้าย คือ ประมาณค่าคะแนนความสามารถของนักเรียน มีขั้นตอนดังนี้

1. ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและแบบสอบและการทำหน้าที่ต่างกันเป็นขั้นตามตัวแปรที่ศึกษา การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและแบบสอบและการทำหน้าที่ต่างกันเป็นขั้นตามตัวแปรที่ศึกษาจำนวน 5 ตัวแปร คือ เพศ เศรษฐานะของครอบครัว แหล่งทรัพยากรการศึกษาที่บ้าน ความมั่งคั่งของครอบครัว และการเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา โดยใช้โปรแกรม DIFAS 5.0 (Penfield, 2005)

สถิติที่ใช้ตรวจสอบการทำหน้าที่ต่างกันของแบบสอบ (differential test functioning: DTF) ในแบบสอบที่มีการตรวจให้คะแนนแบบผสมผสาน (mixed format tests) ใช้การตรวจสอบความแปรปรวนอิทธิพลการทำหน้าที่ต่างกันของข้อสอบแบบถ่วงน้ำหนัก ($\hat{\upsilon}_w^2$) โดยเกณฑ์ประเมินขนาดอิทธิพลแบ่งเป็น 3 ช่วงอิทธิพล ได้แก่ อิทธิพลขนาดเล็ก (S = small DIF effect variance), อิทธิพลขนาดกลาง (M = medium DIF effect variance), และอิทธิพลขนาดใหญ่ (L = large DIF effect variance) (Penfield & Algina, 2006) รายละเอียดตามตารางที่ 25

ตารางที่ 25 อิทธิพลการทำหน้าที่ต่างกันของข้อสอบที่ใช้ประเมินแบบสอบที่ตรวจพบการทำหน้าที่ต่างกัน

ขนาดอิทธิพล	ช่วงการตัดสิน
ขนาดเล็ก (S)	$\hat{\upsilon}_w^2 < .07$
ขนาดกลาง (M)	$.07 \le \hat{\upsilon}_w^2 \le .14$
ขนาดใหญ่ (L)	$\hat{\upsilon}_w^2 > .14$

ส่วนสถิติที่ใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับข้อสอบที่ให้คะแนนทวิวิภาค (differential item functioning: DIF) ด้วยวิธีการวิเคราะห์แมนเทล-แฮนส์เซลร่วมกับอัตราส่วนแต้มต่อมาตรฐาน (standardized mantel-haenszel log-odds ratio: LOR Z) เกณฑ์ที่ใช้ในการตัดสินข้อสอบทำหน้าที่ต่างกันพิจารณาจาก LOR Z > 2.0 (เอนเอียงเข้าข้างไปยังกลุ่มอ้างอิง) และ LOR Z < -2.0 (เอนเอียงเข้าข้างไปยังกลุ่มเปรียบเทียบ) (Penfield, 2007a)

สำหรับข้อสอบที่มีการให้คะแนนพหุวิภาคใช้สถิติตรวจสอบตัวประมาณค่าอิทธิพลการทำหน้าที่ต่างกันของข้อสอบลัว-อเกรสตีปกติกับอัตราส่วนแต้มต่อ (standardized Liu-Agresti common log-odds ratio DIF effect estimator: Z(Lor)) เกณฑ์ที่ใช้ในการตัดสินข้อสอบทำหน้าที่ต่างกันพิจารณาจาก Z(Lor) > 2.0 (เอนเอียงเข้าข้างไปยังกลุ่มอ้างอิง) และ Z(Lor) < -2.0 (เอนเอียงเข้าข้างไปยังกลุ่มเปรียบเทียบ) (Penfield, 2007) ร่วมกับตัวตรวจสอบการทำหน้าที่ต่างกันเป็นขั้นด้วยตัวประมาณค่าอิทธิพลดีเอสเอฟกับอัตราส่วนแต้มต่อระดับขั้น (a step-level log-odds ratio DSF effect estimator) แบบแนวคิดสะสม (the cumulative approach: CU-LOR) แบ่งเป็น 3 ช่วงอิทธิพล คือ ขนาดเล็ก (S) ขนาดกลาง (M) และขนาดใหญ่ (L) (Miller, Chahine, & Childs, 2010) รายละเอียดตามตารางที่ 26

ตารางที่ 26 ขนาดอิทธิพลความแปรปรวนการทำหน้าที่ต่างกันเป็นขั้น

ขนาดอิทธิพล	ช่วงการตัดสิน
ขนาดเล็ก (S)	CU-LOR < 0.43
ขนาดกลาง (M)	0.43 ≤ CU-LOR ≤ 0.64
ขนาดใหญ่ (L)	CU-LOR > 0.64

2. พิจารณาการตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันและโครงสร้างข้อสอบภายหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน ส่วนขั้นตอนนี้เป็นการพิจารณาการตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันและโครงสร้างข้อสอบทั้งก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน ดังนั้น ผู้วิจัยจึงได้กำหนดเกณฑ์ที่ใช้ตัดข้อสอบพร้อมกับการวิเคราะห์คุณภาพของแบบสอบหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน ดังนี้

2.1 กำหนดเกณฑ์การตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน ตามตัวแปรที่ศึกษา ดังนี้ 1) ข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันมาจากตัวแปรนักเรียนมากกว่าหนึ่งตัวแปร 2) ตัดข้อสอบที่ตรวจพบทำหน้าที่ต่างกันบางข้อออกไปแล้วไม่ควรเกินร้อยละ 20 (Clauser, 1993 cited in Naraya & Swaminathan, 1994) ของข้อสอบที่มีอยู่ในแต่ละฉบับ โดยเงื่อนไขการตัดข้อสอบสำหรับข้อสอบที่มีการให้คะแนนทวิวิภาคพิจารณาค่าสถิติแมนเทล-แฮนส์เซลร่วมกับ

อัตราส่วนแต้มต่อมาตรฐาน (standardized mantel-haenszel log-odds ratio: LOR Z) ส่วนข้อสอบที่มีการให้คะแนนพหุวิภาคพิจารณาค่าสถิติตัวประมาณค่าอิทธิพลการทำหน้าที่ต่างกันของข้อสอบลัว-อเกรสตีปกติกับอัตราส่วนแต้มต่อ (standardized Liu-Agresti common log-odds ratio DIF effect estimator: Z(Lor)) ร่วมกับขนาดอิทธิพลการทำหน้าที่ต่างกันเป็นขั้น ขนาดกลาง (M) และขนาดใหญ่ (L) ส่วนแบบสอบที่ตรวจพบทำหน้าที่ต่างกันพิจารณาจากขนาดอิทธิพลขนาดกลาง (M) และขนาดใหญ่ (L) และ 3) ข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันที่ถูกตัดไปควรทำให้อิทธิพลการทำหน้าที่ต่างกันของแบบสอบลดลงหรือเท่าเดิมได้ซึ่งต้องตรวจสอบความตรงเชิงโครงสร้างแบบสอบภายหลังการตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันออกไปแล้ว

2.2 วิเคราะห์คุณภาพของแบบสอบหลังตัดข้อสอบด้วยการคำนวณค่าความเที่ยงแบบความสอดคล้องภายในของแบบสอบและวิเคราะห์ความตรงเชิงโครงสร้างของแบบสอบด้วยการวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA) ตามสมรรถนะทางวิทยาศาสตร์ ด้วยโปรแกรม LISREL 8.72

3. ประมาณค่าคะแนนความสามารถของนักเรียน (theta scores) จากโครงการประเมินผลนักเรียนระดับนานาชาติในวิชาการรู้เรื่องวิทยาศาสตร์ ปี 2009 โดยใช้โปรแกรม IRTPRO 2.1 จำนวน 2 ครั้ง คือ ครั้งที่ 1 ประมาณค่าคะแนนความสามารถของผู้สอบก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน และครั้งที่ 2 ประมาณค่าคะแนนความสามารถของผู้สอบหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน

ทั้งนี้ผลจากการที่นักเรียนทำแบบสอบต่างชุดกันทำให้คะแนนของผู้สอบแต่ละคนได้รับอิทธิพลจากแบบสอบ อันเนื่องมาจากแบบสอบแต่ละฉบับมีค่าความยากแตกต่างกัน ดังนั้น ผู้วิจัยจึงดำเนินการแปลงคะแนนสเกลการประเมินการรู้เรื่องวิทยาศาสตร์ จากการประมาณค่าคะแนนความสามารถของนักเรียน 2 ครั้ง คือ ก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันตามสมการแปลงคะแนนสเกลจากโครงการประเมินผลนักเรียนระดับนานาชาติการรู้เรื่องวิทยาศาสตร์ ปี 2009 (OECD, 2012) ให้อยู่ในคะแนนมาตรฐาน ตามสมการ 3.1 ดังนี้

คะแนนสเกลของ PISA = $((\theta - \bar{x}) / SD) * 100 + 500$ .....(3.1)

โดยที่

θ หมายถึง ค่าประมาณความสามารถของนักเรียน (theta scores)

$\bar{x}$ หมายถึง ค่าเฉลี่ย เท่ากับ .1646

SD หมายถึง ส่วนเบี่ยงเบนมาตรฐาน เท่ากับ 1.0724

5.4 ตรวจสอบเงื่อนไขการวิเคราะห์พหุระดับและความเป็นเอกมิติของแบบสอบและการวิเคราะห์พหุระดับใช้ใช้โมเดลมูลค่าเพิ่มพหุระดับ ในส่วนนี้เป็นตรวจสอบเงื่อนไขการวิเคราะห์พหุระดับและความเป็นเอกมิติของแบบสอบเพื่อนำไปสู่การวิเคราะห์โมเดลมูลค่าเพิ่มพหุระดับ มีขั้นตอนดังนี้

1. การตรวจสอบเงื่อนไขการวิเคราะห์พหุระดับในเบื้องต้น คือการตรวจสอบความสัมพันธ์ระหว่างตัวแปรโดยการตรวจสอบภาวะร่วมเส้นตรงพหุ (Multicollinearity) เป็นการตรวจสอบความสัมพันธ์ระหว่างตัวแปรต้น คู่ที่มีความสัมพันธ์กันสูงหรือภาวะร่วมเส้นตรงพหุ ถ้าหากพบตัวแปรต้นคู่ที่มีความสัมพันธ์กันสูงจะเลือกตัวแปรใดตัวแปรหนึ่งเข้าสู่สมการการวิเคราะห์แทนและการตรวจสอบความสอดคล้องกับสมการการวิเคราะห์ถดถอย (multiple regression) โดยพิจารณาค่า Tolerance และค่า VIF ซึ่งค่า Tolerance ควรสูงกว่า 0.1 และค่า VIF ควรมีค่าน้อยกว่า 10 (Hair et al., 1998)

2. วิเคราะห์คะแนนมูลค่าเพิ่ม (value-added scores) โดยการประยุกต์ใช้โมเดลเชิงเส้นแบบลดหลั่นด้วยการวิเคราะห์ 2 ระดับ ตามโมเดลประเมินคุณภาพการจัดการศึกษาวิชาวิทยาศาสตร์ 2 โมเดล ได้แก่ โมเดล Undetected DIF-DSF-DTF หมายถึง โมเดลการประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ที่ไม่ได้พิจารณาข้อสอบที่ทำหน้าที่ต่างกันข้อสอบและการทำหน้าที่ต่างกันของแบบสอบแต่มีการควบคุมอิทธิพลของตัวแปรในระดับนักเรียนและสถานศึกษา (โมเดลที่ 1) และโมเดล Detected DIF-DSF-DTF หมายถึง โมเดลการประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ที่พิจารณาข้อสอบที่ทำหน้าที่ต่างกันข้อสอบและแบบสอบและมีการควบคุมตัวแปรทั้งในระดับนักเรียนและสถานศึกษา (โมเดลที่ 2) ด้วยโปรแกรม HLM 7.01 สำหรับ การวิเคราะห์มูลค่าเพิ่มในโมเดลการประเมินคุณภาพการจัดการศึกษาทั้ง 2 โมเดล มีสมการวิเคราะห์พหุระดับ 2 ระดับ ดังนี้

ระดับที่ 1 ระดับนักเรียน เป็นการวิเคราะห์ข้อมูลตัวแปรระหว่างนักเรียนภายในหน่วยงานของสถานศึกษา (ระหว่างนักเรียน i ภายในสถานศึกษา j) มีสมการดังนี้

$$Y_{ij} = B_{0j} + \sum_{p=1}^{m} B_{pj}X_{pij} + R_{ij}$$

โดยที่

Y_{ij} คือ ค่าความสามารถจากการประเมินการรู้เรื่องวิทยาศาสตร์ของนักเรียน i ภายในสถานศึกษา j

B_{0j} คือ ค่าจุดตัดแกน (intercept) หรือค่าเฉลี่ยของ Y_{ij} เมื่อควบคุมผลของ X_{pij} แล้ว ซึ่งเป็นค่าที่แสดงถึงคะแนนเฉลี่ยของค่าความสามารถของนักเรียนจากการประเมินการรู้เรื่องวิทยาศาสตร์ของนักเรียนภายในสถานศึกษา j

B_{pj} คือ ค่าความชัน (Slope) หรือค่าสัมประสิทธิ์ถดถอยที่แสดงผลของ X_{pij} ต่อ Y_{ij} ซึ่งเป็นค่าที่แสดงอิทธิพลของการเปลี่ยนแปลงของค่าความสามารถของนักเรียนจากการประเมินการรู้เรื่องวิทยาศาสตร์เมื่อควบคุมตัวแปรของนักเรียน X_{pij}

X_{pij} คือ ตัวแปรควบคุมระดับบุคคล เมื่อ p=1,2,3,..m

R_{ij} คือ ค่าส่วนที่เหลือหรืออิทธิพลสุ่มระดับบุคคลโดยค่า B_{0j} และ B_{pj} ที่ได้จากการวิเคราะห์ระดับที่ 1 จะใช้เป็นตัวแปรตามในการวิเคราะห์ระดับที่ 2

ระดับที่ 2 ระดับสถานศึกษา เป็นการวิเคราะห์ข้อมูลของตัวแปรระหว่างสถานศึกษา มีสมการดังนี้

$$B_{0j} = G_{00} + \sum_{q=1}^{n} G_{0q}Z_j + U_{0j}$$

$$B_{pj} = G_{p0} + \sum_{q=1}^{n} G_{0q}Z_j + U_{pj}$$

โดยที่

B_{0j} คือ ค่าเฉลี่ยของ Y_{ij} เมื่อควบคุมผลของ X_{pij} แล้วซึ่งเป็นค่าที่แสดงถึงคะแนนเฉลี่ยของค่าความสามารถของนักเรียนจากการประเมินการรู้เรื่องวิทยาศาสตร์ของนักเรียนภายในสถานศึกษา j

G_{00} คือ ค่าจุดตัดแกน (intercept) หรือค่าเฉลี่ยของ B_{0j} เมื่อควบคุมผลของ Z_j แล้วซึ่งเป็นค่าแสดงถึงคะแนนเฉลี่ยของค่าความสามารถของนักเรียนจากการประเมินการรู้เรื่องวิทยาศาสตร์ในสถานศึกษาทุกแห่ง

G_{0q} คือ ค่าความชัน (Slope) หรือค่าสัมประสิทธิ์ถดถอยที่แสดงผลของ Z_j ต่อ B_{0j} ซึ่งแสดงถึงค่าอิทธิพลของการเปลี่ยนแปลงของค่าความสามารถของนักเรียนจากการประเมินการรู้เรื่องวิทยาศาสตร์เมื่อมีการควบคุมตัวแปรของสถานศึกษา Y_j

Z_j คือ ตัวแปรควบคุมระดับหน่วยงาน

U_{0j} คือ ค่าส่วนที่เหลือ (residual terms) หรือมูลค่าเพิ่ม (value added) ของสถานศึกษา j

สรุปสมการวิเคราะห์มูลค่าเพิ่มโมเดลเชิงเส้นแบบลดหลั่นด้วยการวิเคราะห์ 2 ระดับ จากทั้ง 2 โมเดล รายละเอียดตามตารางที่ 27

ตารางที่ 27 สมการวิเคราะห์มูลค่าเพิ่มพหุระดับตามโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ของสถานศึกษาทั้ง 2 โมเดล

ระดับการวิเคราะห์	Undetected DIF-DSF-DTF Model	Detected DIF-DSF-DTF Model
ระดับที่ 1	ตัวแปรนักเรียน: เพศ, การเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา, เศรษฐานะของครอบครัว, แหล่งทรัพยากรการศึกษาที่บ้าน, ความมั่งคั่งของครอบครัว, และแหล่งทรัพยากรทางไอซีที ตัวแปรตาม: ค่าความสามารถจากการประเมินการรู้เรื่องวิทยาศาสตร์โดยไม่พิจารณา DIF และ DTF สมการวิเคราะห์ $Y_{ij}=B_{0j}+B_{1j}*(X_{ij})_{ij}+R_{ij}$	ตัวแปรนักเรียน: เพศ, การเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา, เศรษฐานะของครอบครัว, แหล่งทรัพยากรการศึกษาที่บ้าน, ความมั่งคั่งของครอบครัว, และแหล่งทรัพยากรทางไอซีที ตัวแปรตาม: ค่าความสามารถจากการประเมินการรู้เรื่องวิทยาศาสตร์โดยพิจารณา DIF และ DTF สมการวิเคราะห์ $Y_{ij}=B_{0j}+B_{1j}*(X_{ij})_{ij}+R_{ij}$
ระดับที่ 2	ตัวแปรสถานศึกษา: ประเภทสถานศึกษา ระดับการขาดแคลนครู ร้อยละงบประมาณของรัฐที่จัดสรรให้สถานศึกษา ทรัพยากรการเรียน สัดส่วนนักเรียนต่อครู และหลักสูตรการวัดและประเมินผล ตัวแปรตาม: B_{0j} B_{pj} สมการวิเคราะห์: $B_{0j}=G_{00}+G_{01}*(Y_j)+U_{0j}$ $B_1=G_{10}$ $B_2=G_{20}$ $B_3=G_{30}$ . . . $B_{pj}=G_{p0}$ คะแนนมูลค่าเพิ่ม: U_{0j}	ตัวแปรสถานศึกษา: ประเภทสถานศึกษา ระดับการขาดแคลนครู ร้อยละของรัฐที่จัดสรรให้สถานศึกษา ทรัพยากรการเรียน สัดส่วนนักเรียนต่อครู และหลักสูตรการวัดและประเมินผล ตัวแปรตาม: B_{0j} B_{pj} สมการวิเคราะห์: $B_{0j}=G_{00}+G_{01}*(Y_j)+U_{0j}$ $B_1=G_{10}$ $B_2=G_{20}$ $B_3=G_{30}$ . . . $B_{pj}=G_{p0}$ คะแนนมูลค่าเพิ่ม: U_{0j}

5.5 เปรียบเทียบประสิทธิภาพโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ ในส่วนนี้เป็นการเปรียบเทียบประสิทธิภาพโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ของ 2 โมเดล ผู้วิจัยแบ่งออกเป็น 4 ตอนย่อย คือ การวิเคราะห์โมเดลมูลค่าเพิ่มพหุระดับ จากนั้นจึงพัฒนาเกณฑ์การตัดสินการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม ศึกษาความสัมพันธ์ระหว่างโมเดลมูลค่าเพิ่มพหุระดับทั้ง 2 โมเดล และเปรียบเทียบความสอดคล้องของประสิทธิผลของโรงเรียน การจัดอันดับคุณภาพ และการจัดกลุ่มคุณภาพระหว่างโมเดลมูลค่าเพิ่มพหุระดับทั้ง 2 โมเดล มีขั้นตอนดังนี้

1. การวิเคราะห์โมเดลมูลค่าเพิ่มพหุระดับ ซึ่งใช้คะแนนมูลค่าเพิ่ม (value-added scores) จากโมเดลการประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ของสถานศึกษา ตามแนวคิดการประมาณค่าแบบเบส์ (baye's estimation) ซึ่งเป็นค่าเศษเหลือที่ได้จากการวิเคราะห์ระดับที่ 2 (U_{0j}) และคำนวณสหสัมพันธ์ภายในชั้น (Intra-class correlation) และสัมประสิทธิ์การทำนายหรืออธิบายได้ (R^2) ระหว่างโมเดล Undetected DIF-DSF-DTF กับโมเดล Detected DIF-DSF-DTF

2. พัฒนาเกณฑ์การตัดสินการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม (value-added scores rating: VAS rating) ของทั้ง 2 โมเดล จำนวน 5 ระดับ แบ่งเป็น ระดับ A (ดีมาก), ระดับ B (ดี), ระดับ C (พอใช้), ระดับ D (ผ่าน), และระดับ E (ควรปรับปรุง) โดยคำอธิบายระดับการตัดสินการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม รายละเอียดตามตารางที่ 28

ตารางที่ 28 ช่วงคะแนนการตัดสินการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม (VAS rating)

ระดับคุณภาพ	คำอธิบาย	ช่วงการตัดสิน
A	**ดีมาก**	$\bar{x}$ + 2.5SD ขึ้นไป
B	**ดี**	$\bar{x}$+1.5SD - $\bar{x}$+2.5SD
C	**พอใช้**	$\bar{x}$+0.5SD - $\bar{x}$+1.5SD
D	**ผ่าน**	$\bar{x}$ ± 0.5SD
E	**ควรปรับปรุง**	น้อยกว่า $\bar{x}$- 0.5SD

3. ศึกษาความสัมพันธ์ระหว่างโมเดลมูลค่าเพิ่มพหุระดับทั้ง 2 โมเดล

3.1 วิเคราะห์ความสัมพันธ์ระหว่างคะแนนเฉลี่ยการรู้เรื่องวิทยาศาสตร์, คะแนนมูลค่าเพิ่ม (value-added scores: VAS) การจัดอันดับคุณภาพคะแนนมูลค่าเพิ่ม (value-added scores ranking: VAS ranking) และการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม (value-added scores rating: VAS rating)

3.2 วิเคราะห์ความสัมพันธ์ระหว่างคะแนนค่าประมาณความสามารถ (theta scores) และคะแนนมูลค่าเพิ่ม (value-added scores) ตามตัวแปรควบคุมทั้งก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน

4. เปรียบเทียบความสอดคล้องของประสิทธิผลของโรงเรียน การจัดอันดับคุณภาพและการจัดกลุ่มคุณภาพระหว่างโมเดลมูลค่าเพิ่มพหุระดับทั้ง 2 โมเดล

4.1 การเปรียบเทียบความสอดคล้องประสิทธิผลของโรงเรียนจากโมเดลมูลค่าเพิ่มพหุระดับของ 2 โมเดล จากกลุ่มโรงเรียนที่มีประสิทธิผล (effective school) และกลุ่มโรงเรียนที่ไม่มีประสิทธิผล (ineffective school) ซึ่งเป็นคะแนนที่ได้จากการแปลงคะแนนมูลค่าเพิ่มตามเกณฑ์ คือ โรงเรียนที่มีคะแนนมูลค่าเพิ่มตั้งแต่ 0 ขึ้นไป เป็นโรงเรียนที่มีประสิทธิผล ส่วนโรงเรียนที่มีคะแนนมูลค่าเพิ่มน้อยกว่า 0 หรือติดลบ เป็นโรงเรียนที่ไม่มีประสิทธิผล

4.1 เปรียบเทียบความสอดคล้องการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่ม (VAS ranking) ใช้โดยสถิติทดสอบวิลคอกซอน (wilcoxon signed ranks test)

4.2 เปรียบเทียบความสอดคล้องการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม (VAS rating) โดยอาศัยหลักเกณฑ์การวัดความสอดคล้อง (measure of agreement) และการแปลความหมาย ของ Cohen's Kappa ซึ่งใช้ค่าสัมประสิทธิ์ของ kappa (Cohen, 1960) ระหว่างโมเดล Undetected DIF-DSF-DTF กับโมเดล Detected DIF-DSF-DTF รายละเอียดช่วงการตัดสินความสอดคล้องค่าสัมประสิทธิ์ของ kappa รายละเอียดตามตารางที่ 29

ตารางที่ 29 ช่วงการตัดสินความสอดคล้องค่าสัมประสิทธิ์ของ kappa

ช่วงการตัดสิน	การตัดสิน
< 0	less than chance agreement (น้อยมาก)
0.01 – 0.20	slight agreement (น้อย)
0.21 – 0.40	fair agreement (พอใช้)
0.41 – 0.60	moderate agreement (ปานกลาง)
0.61 – 0.80	substantial agreement (ดี)
0.81 – 0.99	almost perfect agreement (ดีมาก/ค่อนข้างสมบูรณ์)

บทที่ 4
ผลการวิเคราะห์ข้อมูล

ในการวิจัยครั้งนี้เป็นการศึกษาผลการทำหน้าที่ต่างกันของข้อสอบ (DIF) ที่มีการตรวจให้คะแนนทวิวิภาคและพหุวิภาค และการทำหน้าที่ต่างกันของแบบทดสอบ (DTF) ในแบบทดสอบการประเมินการรู้เรื่องวิทยาศาสตร์ และเปรียบเทียบประสิทธิภาพของโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์เมื่อใช้โมเดลมูลค่าเพิ่มพหุระดับ 2 โมเดล คือ โมเดลที่ไม่ได้พิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ (โมเดลที่ 1) กับโมเดลที่มีการพิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบโดยการตัดข้อสอบที่มีการทำหน้าที่ต่างกันของข้อสอบและการทำหน้าที่ต่างกันเป็นขั้นที่แตกต่างกัน (โมเดลที่ 2) ซึ่งในการเปรียบเทียบประสิทธิภาพของโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ มีวัตถุประสงค์ย่อยคือ เพื่อศึกษาความสัมพันธ์ระหว่างคะแนนมูลค่าเพิ่มกับ 3 ตัวแปร คือ เศรษฐานะของครอบครัว แหล่งทรัพยากรการศึกษาที่บ้าน และความมั่งคั่งของครอบครัว ทั้งก่อนและหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน และเพื่อเปรียบเทียบความสอดคล้องของการจัดอันดับคุณภาพและการจัดกลุ่มคุณภาพของโมเดลมูลค่าเพิ่มพหุระดับทั้ง 2 โมเดล

นอกจากนี้แบบสอบที่มีการตรวจให้คะแนนทวิวิภาคใช้การตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันด้วย วิธีการวิเคราะห์แมนเทล-แฮนส์เซลร่วมกับอัตราส่วนแต้มต่อมาตรฐาน (standardized mantel-haenszel log-odds ratio: LOR Z) สำหรับข้อสอบที่มีการให้คะแนนพหุวิภาคใช้การตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันโดย สถิติตรวจสอบทำหน้าที่ต่างกันเป็นขั้นด้วยตัวประมาณค่าอิทธิพลดีเอสเอฟกับอัตราส่วนแต้มต่อระดับขั้น (a step-level log-odds ratio DSF effect estimator) แบบแนวคิดสะสม (the cumulative approach: CU-LOR) ร่วมกับตัวประมาณค่าอิทธิพลการทำหน้าที่ต่างกันของข้อสอบอัตราส่วนแต้มต่อลัว-อเกรสตีปกติที่เป็นมาตรฐาน (standardized Liu-Agresti common log-odds ratio DIF effect estimator: Z(Lor)) สำหรับการศึกษาการทำหน้าที่ต่างกันของแบบสอบในแบบสอบที่มีการตรวจให้คะแนนแบบผสมผสาน (mixed format tests) ใช้การตรวจสอบความแปรปรวนอิทธิพลการทำหน้าที่ต่างกันของข้อสอบ (the DIF effect variance) โดยโปรแกรม DIFAS 5.0 (Penfield, 2005) หลังจากนั้นประมาณค่าความสามารถของผู้สอบโดยใช้โปรแกรม IRTPRO 2.1 และวิเคราะห์มูลค่าเพิ่มโดยใช้โมเดลการวิเคราะห์พหุระดับโดยโปรแกรม HLM 7.01

ผู้วิจัยได้นำเสนอผลการวิเคราะห์ข้อมูลเพื่อตอบวัตถุประสงค์การวิจัย โดยแบ่งการนำเสนอผลการวิเคราะห์ข้อมูลเป็น 4 ตอน ดังนี้

ตอนที่ 1 ผลการวิเคราะห์ข้อมูลพื้นฐาน

ตอนที่ 2 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและแบบสอบและการทำหน้าที่ต่างกันเป็นขั้น

ตอนที่ 3 ผลการตรวจสอบเงื่อนไขการวิเคราะห์พหุระดับและความเป็นเอกมิติของแบบสอบและการวิเคราะห์พหุระดับโดยใช้โมเดลมูลค่าเพิ่มพหุระดับ

ตอนที่ 4 ผลการเปรียบเทียบประสิทธิภาพโมเดลการประเมินคุณภาพการจัดการศึกษาวิชาวิทยาศาสตร์

ตอนที่ 1 ผลการวิเคราะห์ข้อมูลพื้นฐาน

ผลการวิเคราะห์ข้อมูลพื้นฐาน ประกอบด้วย ข้อมูลพื้นฐานของกลุ่มตัวอย่างตามโครงการประเมินผลนักเรียนนานาชาติ (PISA) ปี ค.ศ. 2009 ตามตัวแปรระดับนักเรียนและสถานศึกษาเพื่อนำเสนอการกระจายข้อมูลโดยใช้สถิติเชิงบรรยาย ได้แก่ ความถี่ ร้อยละ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ช่วงระหว่างค่าสูงสุดและต่ำสุด ค่าต่ำสุด ค่าสูงสุด ค่าความเบ้และค่าความโด่ง

จากการวิเคราะห์ข้อมูลพื้นฐานของกลุ่มตัวอย่าง ซึ่งเป็นตัวแปรจัดประเภทจำแนกตามตัวแปรนักเรียนและสถานศึกษา ในระดับนักเรียน พบว่า นักเรียนที่ทำการศึกษาส่วนใหญ่เป็นนักเรียนหญิง ร้อยละ 56.8 และเป็นนักเรียนชาย ร้อยละ 43.2 การใช้เวลาเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา พบว่า ส่วนใหญ่ไม่มีเวลาเรียน ร้อยละ 25.3 รองลงมาคือใช้เวลาเรียนพิเศษ 2-4 ชั่วโมง 4-6 ชั่วโมง และน้อยกว่า 2 ชั่วโมง ร้อยละ 11.5, 6.6, และ 6.4 ตามลำดับ และน้อยที่สุดคือใช้เวลาเรียนพิเศษมากกว่า 6 ชั่วโมง ร้อยละ 5.3

ระดับสถานศึกษา พบว่า สถานศึกษาส่วนใหญ่มีขนาดเล็ก ร้อยละ 76.5 รองลงมาขนาดกลางและขนาดใหญ่ ร้อยละ 22.2 และ 1.3 ตามลำดับ ซึ่งเป็นสถานศึกษาของรัฐ ร้อยละ 88.7 และเป็นสถานศึกษาของเอกชน ร้อยละ 11.3 ส่วนสังกัดสถานศึกษา พบว่า ส่วนใหญ่เป็นโรงเรียนในสังกัดกรมสามัญศึกษาเดิม (สพฐ.2) ร้อยละ 39.6 รองลงมาเป็นโรงเรียนในสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานที่มาจากโรงเรียนขยายโอกาสเดิม (สพฐ.1), สังกัดสำนักการศึกษากรุงเทพมหานคร (กทม.), วิทยาลัยอาชีวศึกษาของรัฐ (อศ.2), สังกัดสำนักประสานและพัฒนาการจัดการศึกษาท้องถิ่น (กศท.), และโรงเรียนสาธิตของมหาวิทยาลัย (สาธิต) ร้อยละ 16.5, 9.6, 8.3, 7.8, และ 7.0 ตามลำดับ และน้อยที่สุดเป็นโรงเรียนในสังกัดสำนักบริหารงานคณะกรรมการส่งเสริมการศึกษาเอกชน (สช.) และโรงเรียนอาชีวศึกษาของเอกชน (อศ.1) ร้อยละ 5.7 ตำแหน่งที่ตั้งของสถานศึกษา พบว่า ส่วนใหญ่ตั้งอยู่ในเมือง ร้อยละ 31.7 รองลงมาตั้งอยู่ในหมู่บ้าน เมืองเล็ก และ

เมืองใหญ่ ร้อยละ 23.0, 20.0, และ 16.5 ตามลำดับ และน้อยที่สุดเป็นสถานศึกษาตั้งอยู่ในเมืองใหญ่มาก ร้อยละ 8.7 และเมื่อพิจารณาระดับการขาดแคลนครูวิทยาศาสตร์ในโรงเรียน พบว่าโรงเรียนส่วนใหญ่ไม่ขาดแคลนครูวิทยาศาสตร์ ร้อยละ 39.1 รองลงมาคือขาดแคลนน้อยและขาดแคลนเป็นบางส่วน ร้อยละ 36.1 และ 18.3 ตามลำดับ น้อยที่สุดคือขาดแคลนจำนวนมาก ร้อยละ 6.5 รายละเอียดตามตารางที่ 30

ตารางที่ 30 ความถี่และร้อยละของตัวแปรจัดประเภทจำแนกตามตัวแปรระดับนักเรียนและสถานศึกษา

ตัวแปรระดับนักเรียนและสถานศึกษา	ความถี่	ร้อยละ
ระดับนักเรียน		
1. เพศ		
1.1 ชาย	1,853	43.2
1.2 หญิง	2,439	56.8
2. การใช้เวลาเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา		
2.1 ไม่มีเวลา	1,085	25.3
2.2 น้อยกว่า 2 ชม.	276	6.4
2.3 2-4 ชม.	494	11.5
2.4 4-6 ชม.	283	6.6
2.5 มากกว่า 6 ชม.	226	5.3
ระดับสถานศึกษา		
1. ขนาดสถานศึกษา		
1.1 ขนาดเล็ก (นักเรียน 1- 499 คน)	176	76.5
1.2 ขนาดกลาง (นักเรียน 500 – 1,500 คน)	51	22.2
1.3 ขนาดใหญ่ (นักเรียนมากกว่า 1,500 คน)	3	1.3
2. ประเภทสถานศึกษา		
2.1 สถานศึกษาของรัฐ	204	88.7
2.2 สถานศึกษาของเอกชน	26	11.3

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 30 ความถี่และร้อยละของตัวแปรจัดประเภทจำแนกตามตัวแปรระดับนักเรียนและสถานศึกษา (ต่อ)

ตัวแปรระดับนักเรียนและสถานศึกษา	ความถี่	ร้อยละ
3. สังกัดของสถานศึกษา		
3.1 โรงเรียนในสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ที่มาจากโรงเรียนขยายโอกาสเดิม (สพฐ.1)	38	16.5
3.2 โรงเรียนในสังกัดกรมสามัญศึกษาเดิม (สพฐ.2)	91	39.6
3.3 โรงเรียนในสังกัดสำนักบริหารงานคณะกรรมการส่งเสริมการศึกษาเอกชน(สช.)	13	5.7
3.4 โรงเรียนในสังกัดสำนักการศึกษากรุงเทพมหานคร (กทม.)	22	9.6
3.5 โรงเรียนในสังกัดสำนักประสานและพัฒนาการจัดการศึกษาท้องถิ่น (กศท.)	18	7.8
3.6 โรงเรียนสาธิตของมหาวิทยาลัย (สาธิต)	16	7.0
3.7 โรงเรียนอาชีวศึกษาของเอกชน (อศ.1)	13	5.7
3.8 วิทยาลัยอาชีวศึกษาของรัฐ (อศ.2)	19	8.3
4. ตำแหน่งที่ตั้งของสถานศึกษา		
4.1 หมู่บ้าน (ประชาชนน้อยกว่า 3,000 คน)	53	23.0
4.2 เมืองเล็ก (ประชาชน 3,000-ประมาณ 15,000 คน)	46	20.0
4.3 เมือง (ประชาชน 15,000 -100,000 คน)	73	31.7
4.4 เมืองใหญ่ (ประชาชน 100,000 -1,000,000 คน)	38	16.5
4.5 เมืองใหญ่มาก (ประชาชนมากกว่า 1,000,000 คน)	20	8.7
5. ระดับการขาดแคลนครูวิทยาศาสตร์		
5.1 ไม่ขาดแคลน	90	39.1
5.2 ขาดแคลนน้อย	42	18.3
5.3 ขาดแคลนเป็นบางส่วน	83	36.1
5.4 ขาดแคลนจำนวนมาก	15	6.5

ผลการวิเคราะห์คุณลักษณะพื้นฐานของกลุ่มตัวอย่างที่เป็นตัวแปรต่อเนื่อง จำแนกเป็นระดับนักเรียนและสถานศึกษา ในระดับนักเรียน พบว่า เศรษฐานะของครอบครัวมีค่าเฉลี่ยเท่ากับ -1.16 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.26 ค่าต่ำสุดเท่ากับ -3.90 และค่าสูงสุดเท่ากับ 2.60 แหล่งทรัพยากรการศึกษาที่บ้านมีค่าเฉลี่ยเท่ากับ -0.23 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.08 ค่าต่ำสุดเท่ากับ -3.20 และค่าสูงสุดเท่ากับ 1.36 ความมั่งคั่งของครอบครัวมีค่าเฉลี่ยเท่ากับ -1.16 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.00 ค่าต่ำสุดเท่ากับ -5.08 และค่าสูงสุดเท่ากับ 2.12 และแหล่งทรัพยากรทางไอซีทีมีค่าเฉลี่ยเท่ากับ -.83 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.45 ค่าต่ำสุด

เท่ากับ -2.30 และค่าสูงสุดเท่ากับ 2.49 โดยที่เศรษฐานะของครอบครัว (Sk=0.46, Ku=-0.86) แหล่งทรัพยากรการศึกษาที่บ้าน (Sk=0.01, Ku=-0.82) ความมั่งคั่งของครอบครัว (Sk=0.22, Ku=-0.04) และแหล่งทรัพยากรทางไอซีที (Sk=0.63, Ku=-0.69) มีการกระจายข้อมูลน้อยกว่าค่าเฉลี่ยและต่ำกว่าโค้งปกติ

ระดับสถานศึกษา พบว่า ร้อยละงบประมาณของรัฐที่จัดสรรให้มีค่าเฉลี่ยเท่ากับ 81.91 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 24.24 ค่าต่ำสุดเท่ากับ 0.00 และค่าสูงสุดเท่ากับ 100.00 ทรัพยากรการเรียนมีค่าเฉลี่ยเท่ากับ -0.41 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.01 ค่าต่ำสุดเท่ากับ -2.65 และค่าสูงสุดเท่ากับ 1.93 สัดส่วนนักเรียนต่อครูมีค่าเฉลี่ยเท่ากับ 21.42 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 7.09 ค่าต่ำสุดเท่ากับ 5.05 และค่าสูงสุดเท่ากับ 48.94 และหลักสูตรการวัดและประเมินผลมีค่าเฉลี่ยเท่ากับ 0.81 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.83 ค่าต่ำสุดเท่ากับ -1.14 และค่าสูงสุดเท่ากับ 1.36 โดยที่ร้อยละงบประมาณของรัฐที่จัดสรรให้มากกว่าค่าเฉลี่ยและสูงกว่าโค้งปกติ (Sk=-1.60, Ku=1.92), หลักสูตรการวัดและประเมินผลมากกว่าค่าเฉลี่ยและต่ำกว่าโค้งปกติ (Sk=-0.97, Ku=-0.79), ส่วนทรัพยากรการเรียนในโรงเรียน (Sk=0.75, Ku=0.41) และสัดส่วนนักเรียนต่อครู (Sk=1.03, Ku=2.42) การกระจายข้อมูลน้อยกว่าค่าเฉลี่ยและสูงกว่าโค้งปกติ รายละเอียดตามตารางที่ 31

ตารางที่ 31 ค่าสถิติพื้นฐานของตัวแปรต่อเนื่องจำแนกตามตัวแปรระดับนักเรียนและสถานศึกษา

ตัวแปรระดับนักเรียนและสถานศึกษา	N	Rang	Min	Max	Mean	SD	Sk	Ku
ระดับนักเรียน								
1. เศรษฐานะของครอบครัว	4,281	6.50	-3.90	2.60	-1.16	1.26	0.46	-0.86
2. แหล่งทรัพยากรการศึกษาที่บ้าน	4,286	4.55	-3.20	1.36	-0.23	1.08	0.01	-0.83
3. ความมั่งคั่งของครอบครัว	4,287	7.20	-5.08	2.12	-1.16	1.00	0.22	-0.04
4. แหล่งทรัพยากรทางไอซีที	4,281	4.78	-2.29	2.49	-0.83	1.45	0.63	-0.69
ระดับสถานศึกษา								
1. ร้อยละงบประมาณของรัฐที่จัดสรรให้	230	100.00	0.00	100.00	81.91	24.24	-1.60	1.92
2. ทรัพยากรการเรียน	230	4.58	-2.65	1.93	-0.41	1.01	0.75	0.41
3. สัดส่วนนักเรียนต่อครู	230	43.89	5.05	48.94	21.42	7.09	1.03	2.42
4. หลักสูตรการวัดและประเมินผล	230	2.50	-1.14	1.36	0.81	0.83	-0.97	-0.79

เมื่อพิจารณาคะแนนความสามารถรายบุคคลของนักเรียน (θ) ในวิชาวิทยาศาสตร์จากคะแนนการประเมินผลนักเรียนนานาชาติ (PISA) ปี ค.ศ. 2009 ตามโมเดลการตอบสนองข้อสอบโดยโปรแกรม IRTPRO 2.1 ในระดับนักเรียน พบว่า เพศหญิงมีคะแนนความสามารถเฉลี่ยเท่ากับ 0.05 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.87 ส่วนเพศชายมีคะแนนความสามารถเฉลี่ยเท่ากับ -0.06

และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.89 โดยที่คะแนนความสามารถทั้งเพศหญิง (Sk=0.29, Ku= -0.08) และเพศชาย (Sk=0.33, Ku=-0.37) มีคะแนนความสามารถต่ำกว่าค่าเฉลี่ยและมีการกระจายของคะแนนต่ำกว่าโค้งปกติ **ส่วนการใช้เวลาเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา** พบว่า นักเรียนที่ไม่มีเวลาเรียนพิเศษมีคะแนนความสามารถเฉลี่ยเท่ากับ 0.07 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.83 นักเรียนที่ใช้เวลาเรียนพิเศษน้อยกว่า 2 ชม. มีคะแนนความสามารถเฉลี่ยเท่ากับ 0.12 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.90 นักเรียนที่ใช้เวลาเรียนพิเศษ 2-4 ชม. มีคะแนนความสามารถเฉลี่ยเท่ากับ 0.37 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.96 นักเรียนที่ใช้เวลาเรียนพิเศษ 4-6 ชม. มีคะแนนความสามารถเฉลี่ยเท่ากับ 0.42 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.96 และนักเรียนที่ใช้เวลาเรียนพิเศษมากกว่า 6 ชม. มีคะแนนความสามารถเฉลี่ยเท่ากับ 0.53 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.92 โดยที่คะแนนความสามารถของนักเรียนในกลุ่มใช้เวลาเรียนพิเศษมากกว่า 6 ชม. (Sk=0.08, Ku=-0.60) กลุ่มใช้เวลาเรียนพิเศษน้อยกว่า 2 ชม. (Sk= 0.28, Ku= -0.45) และกลุ่มไม่มีเวลาเรียนเรียนพิเศษ (Sk=0.25, Ku=-0.19) มีคะแนนความสามารถต่ำกว่าค่าเฉลี่ยและมีการกระจายของคะแนนต่ำกว่าโค้งปกติ ส่วนกลุ่มใช้เวลาเรียนพิเศษ 2-4 ชม. (Sk=-0.13, Ku=-0.48) และกลุ่มใช้เวลาเรียนพิเศษ 4-6 ชม. (Sk=-0.09, Ku=-0.49) มีคะแนนค่าความสามารถสูงกว่าค่าเฉลี่ยและมีการกระจายของคะแนนต่ำกว่าโค้งปกติ

ระดับสถานศึกษา พบว่า ขนาดสถานศึกษาขนาดเล็กมีคะแนนความสามารถเฉลี่ยเท่ากับ 0.01 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.84 สถานศึกษาขนาดกลางมีคะแนนความสามารถเฉลี่ยเท่ากับ -0.08 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.93 และขนาดใหญ่มีคะแนนความสามารถเฉลี่ยเท่ากับ 0.18 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.91 โดยที่สถานศึกษาขนาดใหญ่ (Sk=1.73, Ku=0.00) มีคะแนนความสามารถต่ำกว่าค่าเฉลี่ยและมีการกระจายของคะแนนเป็นโค้งปกติ ส่วนสถานศึกษาขนาดเล็ก (Sk=0.41, Ku=-0.46) และขนาดกลาง (Sk=0.54, Ku=-0.28) มีคะแนนความสามารถต่ำกว่าค่าเฉลี่ยและมีการกระจายของคะแนนต่ำกว่าโค้งปกติ

ประเภทของศึกษา พบว่า โรงเรียนรัฐบาลมีคะแนนความสามารถเฉลี่ยเท่ากับ -0.01 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.87 ส่วนโรงเรียนเอกชนมีคะแนนความสามารถเฉลี่ยเท่ากับ 0.01 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.81 โดยที่คะแนนความสามารถของนักเรียนจากโรงเรียนเอกชน (Sk=-0.12, Ku=-0.68) สูงกว่าค่าเฉลี่ยและมีการกระจายของคะแนนต่ำกว่าโค้งปกติ ส่วนคะแนนค่าความสามารถของนักเรียนจากโรงเรียนรัฐบาล (Sk=0.49, Ku=-0.40) ต่ำกว่าค่าเฉลี่ยและมีการกระจายของคะแนนต่ำกว่าโค้งปกติ

สังกัดของสถานศึกษา พบว่า นักเรียนที่เรียนในโรงเรียนสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานที่มาจากโรงเรียนขยายโอกาสเดิม (สพฐ.1) มีคะแนนความสามารถเฉลี่ยเท่ากับ -0.36 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.67 นักเรียนที่เรียนโรงเรียนในสังกัดกรมสามัญศึกษา

เดิม (สพฐ.2) มีคะแนนความสามารถเฉลี่ยเท่ากับ 0.09 และ ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.86 นักเรียนที่เรียนในโรงเรียนสังกัดสำนักบริหารงานคณะกรรมการส่งเสริมการศึกษาเอกชน (สช.) มีคะแนนความสามารถเฉลี่ยเท่ากับ -0.08 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.85 นักเรียนที่เรียนในโรงเรียนสังกัดสำนักการศึกษากรุงเทพมหานคร (กทม.) มีคะแนนความสามารถเฉลี่ยเท่ากับ -0.23 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.79 นักเรียนที่เรียนในโรงเรียนสังกัดสำนักประสานและพัฒนาการจัดการศึกษาท้องถิ่น (กศท.) มีคะแนนความสามารถเฉลี่ยเท่ากับ -0.29 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.72 นักเรียนที่เรียนในโรงเรียนสังกัดโรงเรียนสาธิตของมหาวิทยาลัย (สาธิต) มีคะแนนความสามารถเฉลี่ยเท่ากับ 0.96 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.85 นักเรียนที่เรียนในโรงเรียนสังกัดอาชีวศึกษาของเอกชน (อศ.1) มีคะแนนความสามารถเฉลี่ยเท่ากับ -0.30 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.70 และนักเรียนที่เรียนในโรงเรียนสังกัดวิทยาลัยอาชีวศึกษาของรัฐ (อศ.2) มีคะแนนความสามารถเฉลี่ยเท่ากับ -0.33 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.76 โดยที่นักเรียนที่เรียนอยู่สังกัดโรงเรียนสาธิตของมหาวิทยาลัย (Sk=-0.42, Ku=0.07) มีคะแนนสูงกว่าค่าเฉลี่ยและมีการกระจายของคะแนนใกล้เคียงโค้งปกติ ส่วนโรงเรียนในสังกัดกรมสามัญศึกษาเดิม (สพฐ.2) (Sk=0.19, Ku=-0.25) สังกัดสำนักบริหารงานคณะกรรมการส่งเสริมการศึกษาเอกชน(สช.) (Sk=0.20, Ku=-0.52) สังกัดสำนักการศึกษากรุงเทพมหานคร (กทม.) (Sk=0.24, Ku=-0.16), สังกัดสำนักประสานและพัฒนาการจัดการศึกษาท้องถิ่น (กศท.) (Sk=0.14, Ku=-0.29) โรงเรียนอาชีวศึกษาของเอกชน (อศ.1) (Sk=0.19, Ku=-0.01) และวิทยาลัยอาชีวศึกษาของรัฐ (อศ.2) (Sk=0.29, Ku=-0.30) มีคะแนนต่ำว่าค่าเฉลี่ยและมีการกระจายของคะแนนต่ำกว่าโค้งปกติ ส่วนนักเรียนที่เรียนในโรงเรียนสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานที่มาจากโรงเรียนขยายโอกาสเดิม (สพฐ.1) (Sk=0.31, Ku=0.62) มีคะแนนต่ำกว่าค่าเฉลี่ยและมีการกระจายของคะแนนสูงกว่าโค้งปกติ

ตำแหน่งที่ตั้งของสถานศึกษา พบว่า ตำแหน่งที่ตั้งของสถานศึกษาในหมู่บ้านมีคะแนนความสามารถเฉลี่ยเท่ากับ -0.07 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.87 ตำแหน่งที่ตั้งของสถานศึกษาในเมืองเล็กมีคะแนนความสามารถเฉลี่ยเท่ากับ 0.27 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.92 ตำแหน่งที่ตั้งของสถานศึกษาในเมืองมีคะแนนความสามารถเฉลี่ยเท่ากับ -0.06 และ ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.72 ตำแหน่งที่ตั้งของสถานศึกษาในเมืองใหญ่มีคะแนนความสามารถเฉลี่ยเท่ากับ -0.07 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.93 และตำแหน่งที่ตั้งของสถานศึกษาในเมืองใหญ่มากมีคะแนนความสามารถเฉลี่ยเท่ากับ -0.20 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.97 โดยที่คะแนนความสามารถของนักเรียนที่มีตำแหน่งที่ตั้งของสถานศึกษาอยู่ในหมู่บ้าน (Sk=0.44, Ku=-0.42) เมืองขนาดเล็ก (Sk=0.03, Ku=-1.16) และเมืองใหญ่ (Sk=0.46, Ku=-0.26) มีคะแนนความสามารถต่ำกว่าค่าเฉลี่ยและมีการกระจายของคะแนนต่ำกว่าโค้งปกติ ส่วนคะแนน

ความสามารถของนักเรียนที่มีตำแหน่งที่ตั้งของสถานศึกษาอยู่ในเมือง (Sk=0.44, Ku=0.38) และเมืองใหญ่มาก (Sk=1.15, Ku=0.62) มีคะแนนความสามารถต่ำกว่าค่าเฉลี่ยและมีการกระจายของคะแนนสูงกว่าโค้งปกติ

ระดับการขาดแคลนครูวิทยาศาสตร์ พบว่า โรงเรียนที่ไม่การขาดแคลนครูวิทยาศาสตร์มีคะแนนความสามารถเฉลี่ยเท่ากับ -0.04 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.85 โรงเรียนที่มีการขาดแคลนครูวิทยาศาสตร์ระดับน้อยมีคะแนนความสามารถเฉลี่ยเท่ากับ 0.03 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.87 โรงเรียนที่มีการขาดแคลนครูวิทยาศาสตร์ระดับขาดแคลนเป็นบางส่วนมีคะแนนความสามารถเฉลี่ยเท่ากับ -0.04 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.87 และโรงเรียนที่มีการขาดแคลนครูวิทยาศาสตร์ระดับขาดแคลนจำนวนมากมีคะแนนความสามารถเฉลี่ยเท่ากับ 0.20 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.89 โดยที่คะแนนความสามารถของนักเรียนในโรงเรียนที่ไม่มีการขาดแคลนครูวิทยาศาสตร์ (Sk=0.31, Ku=-0.40), ระดับขาดแคลนน้อย (Sk= 0.59, Ku=-0.68), ระดับขาดแคลนเป็นบางส่วน (Sk=0.56, Ku=-0.10), และระดับขาดแคลนจำนวนมาก (Sk=0.10, Ku=-1.26) มีคะแนนต่ำกว่าค่าเฉลี่ยและมีการกระจายของคะแนนต่ำกว่าโค้งปกติ รายละเอียดตามตารางที่ 32

ตารางที่ 32 ค่าสถิติพื้นฐานคะแนนความสามารถทางวิทยาศาสตร์จำแนกตามตัวแปรระดับนักเรียนและสถานศึกษา

ตัวแปรระดับนักเรียนและสถานศึกษา	N	Range	Min	Max	Mean	SD	Sk	Ku
ระดับนักเรียน								
1. เพศ								
1.1 หญิง	1,853	5.24	-2.21	3.03	0.05	0.87	0.29	-0.08
1.2 ชาย	2,439	4.81	-2.21	2.61	-0.06	0.89	0.33	-0.37
2. การใช้เวลาเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา								
2.1 ไม่มีเวลา	1,085	4.81	-2.03	2.79	0.07	0.83	0.25	-0.19
2.2 น้อยกว่า 2 ชม.	276	4.42	-1.82	2.61	0.12	0.90	0.28	-0.45
2.3 2-4 ชม.	494	4.77	-2.03	2.74	0.37	0.96	-0.13	-0.48
2.4 4-6 ชม.	283	4.60	-1.95	2.65	0.42	0.96	-0.09	-0.49
2.5 มากกว่า 6 ชม.	226	4.48	-1.45	3.03	0.53	0.92	0.08	-0.60

ตารางที่ 32 ค่าสถิติพื้นฐานคะแนนความสามารถทางวิทยาศาสตร์จำแนกตามตัวแปรระดับนักเรียนและสถานศึกษา(ต่อ)

ตัวแปรระดับนักเรียนและสถานศึกษา	N	Range	Min	Max	Mean	SD	Sk	Ku
ระดับสถานศึกษา								
1. ขนาดสถานศึกษา								
1.1 ขนาดเล็ก	176	3.94	-1.88	2.06	0.01	0.84	0.41	-0.46
1.2 ขนาดกลาง	51	3.61	-1.55	2.06	-0.08	0.93	0.54	-0.28
1.2 ขนาดใหญ่	3	1.58	-0.35	1.23	0.18	0.91	1.73	0.00
2. ประเภทสถานศึกษา								
2.1 รัฐบาล	204	3.94	-1.88	2.06	-0.01	0.87	0.49	-0.40
2.2 เอกชน	26	3.07	-1.55	1.52	0.01	0.81	-0.12	-0.68
3. สังกัดของสถานศึกษา								
3.1 สพฐ.1	38	4.27	-2.03	2.24	-0.36	0.67	0.31	0.62
3.2 สพฐ.2	91	4.99	-2.21	2.79	0.09	0.86	0.19	-0.25
3.3 สช.	13	4.04	-1.96	2.08	-0.08	0.85	0.20	-0.52
3.4 กทม.	22	4.30	-2.03	2.27	-0.23	0.79	0.24	-0.16
3.5 กศท.	18	3.64	-1.97	1.68	-0.29	0.72	0.14	-0.29
3.6 สาธิต	16	4.59	-1.56	3.03	0.96	0.85	-0.42	0.07
3.7 อศ.1	13	3.88	-2.21	1.68	-0.30	0.70	0.19	-0.01
3.8 อศ.2	19	4.12	-1.88	2.24	-0.33	0.76	0.29	-0.30
4. ตำแหน่งที่ตั้งของสถานศึกษา								
4.1 หมู่บ้าน	53	3.68	-1.88	1.80	-0.07	0.87	0.44	-0.42
4.2 เมืองเล็ก	46	3.28	-1.22	2.06	0.27	0.92	0.03	-1.16
4.3 เมือง	73	3.36	-1.55	1.80	-0.06	0.72	0.44	0.38
4.4 เมืองใหญ่	38	3.36	-1.55	1.80	-0.07	0.93	0.46	-0.26
4.5 เมืองใหญ่มาก	20	3.28	-1.22	2.06	-0.20	0.97	1.15	0.62
5. ระดับการขาดแคลนครูวิทยาศาสตร์								
5.1 ไม่ขาดแคลน	90	3.94	-1.88	2.06	-0.04	0.85	0.31	-0.40
5.2 ขาดแคลนน้อย	42	3.28	-1.22	2.06	0.03	0.87	0.59	-0.68
5.3 ขาดแคลนเป็นบางส่วน	83	3.36	-1.55	1.80	-0.04	0.87	0.56	-0.10
5.4 ขาดแคลนจำนวนมาก	15	2.74	-1.22	1.52	0.20	0.89	0.10	-1.26

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตอนที่ 2 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ และการทำหน้าที่ต่างกันเป็นขั้น

ในส่วนนี้ผู้วิจัยใช้โปรแกรม DIFAS 5.0 (Penfield, 2005) เข้ามาศึกษาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบและการทำหน้าที่ต่างกันเป็นขั้น เพื่อนำผลที่ได้มาใช้ในการพิจารณาคัดเลือกข้อสอบที่มีความเหมาะสมหลังจากตัดข้อสอบที่เกิดความลำเอียงตามเกณฑ์ที่กำหนดไว้ จากนั้น ทดสอบการทำหน้าที่ต่างกันของแบบสอบอีกครั้งพร้อมกับตรวจสอบโครงสร้างของแบบสอบด้วยการตรวจสอบความตรงเชิงโครงสร้างด้วยการวิเคราะห์องค์ประกอบเชิงยืนยันและนำคะแนนผลการสอบที่ได้ไปคำนวณหาคะแนนความสามารถของผู้สอบอีกครั้ง ดังนั้นผลการวิจัยส่วนนี้ผู้วิจัยจึงแบ่งผลการวิเคราะห์เป็น 3 ขั้นตอนย่อย ได้แก่ ตอนย่อยที่ 2.1 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและแบบสอบและการทำหน้าที่ต่างกันเป็นขั้นตามตัวแปรที่ศึกษา ตอนย่อยที่ 2.2 ผลการพิจารณาการตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันและโครงสร้างข้อสอบทั้งก่อนและภายหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน และตอนย่อยที่ 2.3 ผลการประมาณคะแนนความสามารถของนักเรียนหลังจากตรวจสอบโครงสร้างข้อสอบ

2.1 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ และการทำหน้าที่ต่างกันเป็นขั้นตามตัวแปรที่ศึกษา

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) และแบบสอบ (DTF) และการทำหน้าที่ต่างกันเป็นขั้น (DSF) เป็นการวิเคราะห์หาอิทธิพลของตัวแปรระดับนักเรียนที่มีผลต่อโอกาสการตอบข้อสอบถูกในข้อสอบที่มีการตรวจให้คะแนนทวิวิภาคและพหุวิภาค ผู้วิจัยได้พิจารณาตัวแปรเพื่อนำมาจัดกลุ่มคุณลักษณะของนักเรียนที่เป็นตัวแปรภายนอกซึ่งอยู่นอกเหนือการควบคุมให้เกิดขึ้นได้เอง จึงกำหนดตัวแปรที่นำมาใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ และการทำหน้าที่ต่างกันเป็นขั้นดังต่อไปนี้

GENDER คือ เพศของนักเรียน (นักเรียนหญิง=1, นักเรียนชาย=2)

EXTRA คือ การใช้เวลาเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา (นักเรียนที่เรียนพิเศษ=1, นักเรียนที่ไม่เรียนพิเศษ=2)

ESCS คือ เศรษฐานะของครอบครัว (นักเรียนที่มีเศรษฐานะของครอบครัวสูงกว่าค่าเฉลี่ย=1, นักเรียนที่มีเศรษฐานะของครอบครัวต่ำกว่าค่าเฉลี่ย=2)

HEDRES คือ แหล่งทรัพยากรการศึกษาที่บ้าน (นักเรียนที่มีแหล่งทรัพยากรการศึกษาที่บ้านสูงกว่าค่าเฉลี่ย =1, นักเรียนที่มีแหล่งทรัพยากรการศึกษาที่บ้านต่ำกว่าค่าเฉลี่ย=2)

WEALTH คือ ความมั่งคั่งของครอบครัว (นักเรียนที่มีความมั่งคั่งของครอบครัวสูงกว่าค่าเฉลี่ย=1, นักเรียนที่มีความมั่งคั่งของครอบครัวต่ำกว่าค่าเฉลี่ย =2)

ผู้วิจัยสรุปจำนวนข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันในแต่ละฉบับและนำเสนอรายละเอียดผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ และการทำหน้าที่ต่างกันเป็นขั้นในภาคผนวก ก

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ในข้อสอบที่มีการให้คะแนนทวิวิภาคและพหุวิภาค ซึ่งข้อสอบที่มีการให้คะแนนพหุวิภาคเป็นการศึกษาร่วมกันระหว่างการทำหน้าที่ต่างกันของข้อสอบ (DIF) กับการศึกษาการทำหน้าที่ต่างกันเป็นขั้น (DSF) ทั้งนี้ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจำแนกตามสมรรถนะทางวิทยาศาสตร์ คือ การระบุประเด็นทางวิทยาศาสตร์ (ISI=identify scientific issues) การอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ (EPS=explain phenomena scientifically) และการใช้ประจักษ์พยานทางวิทยาศาสตร์ (USE=using scientific evidence) จาก 5 ตัวแปรที่ศึกษา พบว่า แบบสอบฉบับที่ 7 และ 13 พบว่า มีข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันมากที่สุด ร้อยละ 50 โดยที่แบบสอบฉบับที่ 7 และ 13 ตรวจพบข้อสอบทำหน้าที่ต่างกันแบบตรวจให้คะแนนทวิวิภาค จำนวน 8 ข้อ และแบบตรวจให้คะแนนพหุวิภาค จำนวน 1 ข้อ เป็นจำนวนเท่ากัน รองลงมาคือ แบบสอบฉบับที่ 10 พบว่า มีข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันร้อยละ 48.57 ตรวจพบข้อสอบทำหน้าที่ต่างกันแบบตรวจให้คะแนนทวิวิภาค จำนวน 15 ข้อ และแบบตรวจให้คะแนนพหุวิภาค จำนวน 2 ข้อ แบบสอบฉบับที่ 9 และ 12 พบว่า มีข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันร้อยละ 44.44 โดยที่แบบสอบฉบับที่ 9 ตรวจพบข้อสอบทำหน้าที่ต่างกันแบบตรวจให้คะแนนทวิวิภาค จำนวน 7 ข้อ และแบบตรวจให้คะแนนพหุวิภาค จำนวน 1 ข้อ ส่วนแบบสอบฉบับที่ 12 ตรวจพบข้อสอบทำหน้าที่ต่างกันแบบตรวจให้คะแนนทวิวิภาค จำนวน 14 ข้อ และแบบตรวจให้คะแนนพหุวิภาค จำนวน 2 ข้อ และแบบสอบฉบับที่ 4 พบว่า มีข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันน้อยที่สุด ร้อยละ 16.67 และตรวจพบข้อสอบทำหน้าที่ต่างกันแบบตรวจให้คะแนนทวิวิภาค จำนวน 2 ข้อ และแบบตรวจให้คะแนนพหุวิภาค จำนวน 1 ข้อ รายละเอียดตามตารางที่ 33

ตารางที่ 33 ผลข้อสอบที่ตรวจพบทำหน้าที่ต่างกันรวมทั้ง 5 ตัวแปรในแบบสอบทั้ง 9 ฉบับแยกตามสมรรถนะทางวิทยาศาสตร์

ฉบับที่	สมรรถนะทางวิทยาศาสตร์						รวม	ร้อยละของ DIF
	ให้คะแนนทวิวิภาค (dichotomous items)			ให้คะแนนพหุวิภาค (polytomous items)				
	ISI	EPS	USE	ISI	EPS	USE		
2	6(3)	5	5	-	-	1(1)	17(4)	23.53
3	9(4)	11(3)	13(2)	-	-	2(2)	35(11)	31.43
4	4(1)	11(1)	2	-	-	1(1)	18(3)	16.67
7	3(1)	6(2)	8(5)	-	-	1(1)	18(9)	50.00
8	6(1)	5(2)	5(3)	-	-	1(1)	17(7)	41.18
9	4(2)	11(5)	2	-	-	1(1)	18(8)	44.44
10	10(8)	16(5)	7(2)	-	-	2(2)	35(17)	48.57
12	7(1)	17(8)	10(5)	-	-	2(2)	36(16)	44.44
13	3	6(3)	8(5)	-	-	1(1)	18(9)	50.00

หมายเหตุ ตัวเลขในวงเล็บเป็นจำนวนข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน

เมื่อแยกมิติในการพิจารณาข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันตามตัวแปรของนักเรียนจำนวน 5 ตัวแปร ได้แก่ เพศ การเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา เศรษฐานะของครอบครัว แหล่งทรัพยากรการศึกษาที่บ้าน และความมั่งคั่งของครอบครัว ตามสมรรถนะทางวิทยาศาสตร์ 3 ด้าน คือ การระบุประเด็นทางวิทยาศาสตร์ การอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ และการใช้ประจักษ์พยานทางวิทยาศาสตร์ โดยภาพรวม พบว่า มีข้อสอบเอนเอียงเข้าข้างนักเรียนหญิงมากกว่านักเรียนชายจำนวน 3 ข้อ (ร้อยละ 10.34), เอนเอียงเข้าข้างนักเรียนที่เรียนพิเศษวิทยาศาสตร์นอกสถานศึกษามากกว่านักเรียนที่ไม่เรียนเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา จำนวน 1 ข้อ (ร้อยละ 14.28) เอนเอียงเข้าข้างนักเรียนนักเรียนที่มีเศรษฐานะของครอบครัวต่ำมากกว่านักเรียนนักเรียนที่มีเศรษฐานะของครอบครัวสูง จำนวน 2 ข้อ (ร้อยละ 9.1) เอนเอียงเข้าข้างนักเรียนที่มีแหล่งทรัพยากรการศึกษาที่บ้านมากกว่านักเรียนที่มีแหล่งทรัพยากรการศึกษาที่บ้านต่ำ จำนวน 5 ข้อ (ร้อยละ 29.42) และเอนเอียงเข้าข้างนักเรียนที่มีความมั่งคั่งทางครอบครัวสูงมากกว่านักเรียนที่มีความมั่งคั่งทางครอบครัวต่ำ จำนวน 2 ข้อ (ร้อยละ 16.66)

เมื่อพิจารณาตัวแปรตามสมรรถนะทางวิทยาศาสตร์ที่ศึกษา พบว่า ตัวแปรเพศ ตามสมรรถนะการระบุประเด็นทางวิทยาศาสตร์เอนเอียงเข้าข้างนักเรียนหญิงมากกว่านักเรียนชาย

จำนวน 4 ข้อ (ร้อยละ 13.79) การอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์เอนเอียงเข้าข้างนักเรียนชายมากกว่านักเรียนหญิง จำนวน 1 ข้อ (ร้อยละ 13.79) ส่วนการใช้ประจักษ์พยานทางวิทยาศาสตร์ไม่เอนเอียงเข้าข้างทั้งนักเรียนชายและนักเรียนหญิง

ตัวแปรการเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษาตามสมรรถนะการระบุประเด็นทางวิทยาศาสตร์เอนเอียงเข้าข้างนักเรียนที่ไม่เรียนพิเศษมากกว่านักเรียนที่เรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา จำนวน 1 ข้อ (ร้อยละ 14.28) การใช้ประจักษ์พยานทางวิทยาศาสตร์เอนเอียงเข้าข้างนักเรียนที่เรียนพิเศษมากกว่านักเรียนที่ไม่เรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา จำนวน 2 ข้อ (ร้อยละ 28.57) ส่วนการอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ไม่เอนเอียงเข้าข้างทั้งนักเรียนที่เรียนพิเศษและไม่เรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา

ตัวแปรเศรษฐานะของครอบครัวตามสมรรถนะการระบุประเด็นทางวิทยาศาสตร์เอนเอียงเข้าข้างนักเรียนที่มีเศรษฐานะต่ำมากกว่านักเรียนที่มีเศรษฐานะสูง จำนวน 2 ข้อ (ร้อยละ 9.09) และการอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์เอนเอียงเข้าข้างนักเรียนที่มีเศรษฐานะต่ำมากกว่านักเรียนที่มีเศรษฐานะสูง จำนวน 1 ข้อ (ร้อยละ 4.55) ส่วนการใช้ประจักษ์พยานทางวิทยาศาสตร์เอนเอียงเข้าข้างนักเรียนที่มีเศรษฐานะสูงมากกว่านักเรียนที่มีเศรษฐานะต่ำ จำนวน 1 ข้อ (ร้อยละ 4.55)

ตัวแปรแหล่งทรัพยากรการศึกษาที่บ้านตามสมรรถนะการระบุประเด็นทางวิทยาศาสตร์เอนเอียงเข้าข้างนักเรียนที่มีแหล่งทรัพยากรการศึกษาที่บ้านต่ำมากกว่านักเรียนที่มีแหล่งทรัพยากรการศึกษาที่บ้านสูง จำนวน 2 ข้อ (ร้อยละ 11.77) ส่วนการอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์เอนเอียงเข้าข้างนักเรียนที่มีแหล่งทรัพยากรการศึกษาที่บ้านสูงมากกว่านักเรียนที่มีแหล่งทรัพยากรการศึกษาที่บ้านต่ำ จำนวน 2 ข้อ (ร้อยละ 11.77) ส่วนการใช้ประจักษ์พยานทางวิทยาศาสตร์เอนเอียงเข้าข้างนักเรียนที่มีแหล่งทรัพยากรการศึกษาที่บ้านสูงมากกว่านักเรียนที่มีแหล่งทรัพยากรการศึกษาที่บ้านต่ำ จำนวน 5 ข้อ (ร้อยละ 29.41)

และตัวแปรความมั่งคั่งของครอบครัวตามสมรรถนะการระบุประเด็นทางวิทยาศาสตร์เอนเอียงเข้าข้างนักเรียนที่มีความมั่งคั่งครอบครัวต่ำมากกว่านักเรียนที่มีความมั่งคั่งครอบครัวสูง จำนวน 1 ข้อ (ร้อยละ 8.33), และการอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์เอนเอียงเข้าข้างนักเรียนที่มีความมั่งคั่งครอบครัวต่ำมากกว่านักเรียนที่มีความมั่งคั่งครอบครัวสูง จำนวน 1 ข้อ (ร้อยละ 8.34), ส่วนการใช้ประจักษ์พยานทางวิทยาศาสตร์เอนเอียงเข้าข้างนักเรียนที่มีความมั่งคั่งครอบครัวสูงมากกว่านักเรียนที่มีความมั่งคั่งครอบครัวต่ำ จำนวน 4 ข้อ (ร้อยละ 33.33) รายละเอียดตามตารางที่ 34

ตารางที่ 34 ข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันในแต่ละตัวแปรที่ศึกษา

ตัวแปร		สมรรถนะทางวิทยาศาสตร์			รวม	ร้อยละ
		ISI	EPS	USE		
เพศ	หญิง	6 (20.69%)	5 (17.24%)	5 (17.24%)	16	55.17
	ชาย	2 (6.90%)	6 (20.69%)	5 (17.24%)	13	44.83
การเรียนพิเศษวิทยาศาสตร์	เรียนพิเศษ	1 (14.29%)	0 (0%)	3 (42.86%)	4	57.14
	ไม่เรียนพิเศษ	2 (28.57%)	0 (0%)	1 (14.29%)	3	42.86
เศรษฐานะของครอบครัว	เศรษฐานะสูง	1 (4.55%)	6 (27.27%)	3 (13.64%)	10	45.45
	เศรษฐานะต่ำ	3 (13.64%)	7 (31.82%)	2 (9.09%)	12	54.55
แหล่งทรัพยากรการศึกษาที่บ้าน	ทรัพยากรสูง	1 (5.88%)	4 (23.53%)	6 (35.29%)	11	64.71
	ทรัพยากรต่ำ	3 (17.65%)	2 (11.76%)	1 (5.88%)	6	35.29
ความมั่งคั่งของครอบครัว	ความมั่งคั่งสูง	2 (16.67%)	1 (8.33%)	4 (33.33%)	7	58.33
	ความมั่งคั่งต่ำ	3 (25.00%)	2 (16.67%)	0 (0%)	5	41.67

หมายเหตุ ตัวเลขในวงเล็บเป็นร้อยละข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน และ ISI หมายถึง การระบุประเด็นทางวิทยาศาสตร์, EPS หมายถึง การอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์, USE หมายถึง การใช้ประจักษ์พยานทางวิทยาศาสตร์

ผลการตรวจสอบการทำหน้าที่ต่างกันของแบบสอบ (DTF) ในแบบสอบทั้ง 9 ฉบับ คือ ฉบับที่ 2, 3, 4, 7, 8, 9, 10, 12, และ 13 ในแบบสอบที่ตรวจพบการทำหน้าที่ต่างกันสำหรับแบบสอบที่มีการให้คะแนนแบบผสมผสาน โดยใช้การตรวจสอบความแปรปรวนอิทธิพลการทำหน้าที่ต่างกันของข้อสอบ (the DIF effect variance) แบ่งเป็น 3 ระดับ คือ อิทธิพลขนาดเล็ก (S=small DIF effect variance) อิทธิพลขนาดกลาง (M= medium DIF effect variance) และอิทธิพลขนาดใหญ่ (L=large DIF effect variance)

ทั้งนี้ผู้วิจัยใช้ระดับในการศึกษาอิทธิพลขนาดกลางและขนาดใหญ่เป็นเกณฑ์ในการศึกษาแบบสอบที่ตรวจพบการทำหน้าที่ต่างกัน ซึ่งเกิดจากคุณลักษณะของนักเรียนต่างกลุ่มในตัวแปรเพศ (GENDER) การใช้เวลาเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา (EXTRA) เศรษฐานะของครอบครัว (ESCS) แหล่งทรัพยากรการศึกษาที่บ้าน (HEDRES) และความมั่งคั่งของครอบครัว (WEALTH) พบว่า แบบสอบฉบับที่ 3 พบร้อยละของแบบสอบที่ตรวจพบการทำหน้าที่ต่างกันมากที่สุด ร้อยละ 60 รองลงมาคือแบบสอบฉบับที่ 12 และ 13 ร้อยละ 40 และแบบสอบฉบับที่ 7 และ 10 ร้อยละ 20 นอกจากนี้ในแบบสอบฉบับที่ 2, 4, 8, และ 9 พบแบบสอบที่ตรวจพบการทำหน้าที่ต่างกันในระดับอิทธิพลเล็กสุด ร้อยละ 0

เมื่อพิจารณาตามตัวแปรเศรษฐานะของครอบครัว พบว่า แบบสอบที่ตรวจพบการทำหน้าที่ต่างกันมากที่สุด 4 ฉบับ คือ แบบสอบฉบับที่ 3 7, 10, และ 12 รองลงมาคือตัวแปรความมั่งคั่งของครอบครัว พบว่า แบบสอบที่ตรวจพบการทำหน้าที่ต่างกัน 3 ฉบับ คือ แบบสอบฉบับที่ 3, 12, และ 13 ตัวแปรแหล่งทรัพยากรการศึกษาที่บ้าน พบว่า แบบสอบที่ตรวจพบการทำหน้าที่ต่างกัน 2 ฉบับ คือ แบบสอบฉบับที่ 3 และ 13 ส่วนตัวแปรเพศและตัวแปรการใช้เวลาเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา พบว่า แบบสอบที่ตรวจพบการทำหน้าที่ต่างกันในระดับอิทธิพลเล็กที่สุด รายละเอียดตามตารางที่ 35

ตารางที่ 35 ผลการตรวจสอบการทำหน้าที่ต่างกันในแบบสอบ 9 ฉบับตามตัวแปรที่ศึกษา

ฉบับที่	รูปแบบการทำหน้าที่ต่างกันของแบบสอบตามตัวแปร					ร้อยละของ DTF (พิจารณา M-L)
	GENDER	EXTRA	ESCS	HEDRES	WEALTH	
2	S	S	S	S	S	0%
3	S	S	M	M	M	60%
4	S	S	S	S	S	0%
7	S	S	M	S	S	20%
8	S	S	S	S	S	0%
9	S	S	S	S	S	0%
10	S	S	M	S	S	20%
12	S	S	M	S	M	40%
13	S	S	S	M	M	40%

2.2 ผลการพิจารณาการตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันและโครงสร้างข้อสอบหลังการตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน

ผลการตัดข้อสอบใช้การตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันตามเกณฑ์ที่กำหนดไว้ดังนี้ 1) ข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันมาจากตัวแปรนักเรียนมากกว่าหนึ่งตัวแปร 2) ตัดข้อสอบที่ตรวจพบทำหน้าที่ต่างกันบางข้อออกไปแล้วไม่ควรเกินร้อยละ 20 (Clauser, 1993 cited in Naraya & Swaminathan, 1994) ของข้อสอบที่มีอยู่ในแต่ละฉบับ โดยเงื่อนไขการตัดข้อสอบสำหรับข้อสอบที่มีการให้คะแนนทวิวิภาค พิจารณาค่าสถิติแมนเทล-แฮนส์เซลร่วมกับอัตราส่วนแต้มต่อมาตรฐาน (standardized mantel-haenszel log-odds ratio: LOR Z) ส่วนข้อสอบที่มีการให้คะแนนพหุวิภาค พิจารณาค่าสถิติตัวประมาณค่าอิทธิพลการทำหน้าที่ต่างกันของข้อสอบลัว-อเกรสตีปกติกับอัตราส่วนแต้มต่อ (standardized Liu-Agresti common log-odds ratio DIF effect estimator: Z(Lor)) ร่วมกับขนาดอิทธิพลการทำหน้าที่ต่างกันเป็นขั้น ขนาดกลาง (M) และขนาดใหญ่ (L) ส่วนแบบสอบที่ตรวจพบทำหน้าที่ต่างกันพิจารณาจากขนาดอิทธิพลขนาดกลาง (M) และขนาดใหญ่ (L) และ 3) ข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันที่ถูกตัดไปควรทำให้อิทธิพลการทำหน้าที่ต่างกันของแบบสอบลดลงหรือเท่าเดิมได้ซึ่งต้องตรวจสอบความตรงเชิงโครงสร้างแบบสอบภายหลังการตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันออกไปแล้ว

เมื่อพิจารณาลักษณะข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน ซึ่งถูกตัดตามตัวแปรของนักเรียนโดยจำแนกเป็นข้อสอบที่มีการตรวจให้คะแนนทวิภาคและพหุวิภาค พบว่า แบบสอบฉบับที่ 7, 8, 9, 10, 12, และ 13 มีการตัดข้อสอบออกไปทั้งข้อที่มีการตรวจให้คะแนนทวิภาคและ พหุวิภาค ส่วนแบบสอบฉบับที่ 2, 3, และ 4 มีการตัดข้อสอบออกไปเฉพาะข้อสอบที่มีการตรวจให้คะแนนพหุวิภาค ทั้งนี้แบบสอบฉบับที่ 2 และ 3 ตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันจำนวน 3 ตัวแปร ได้แก่ เศรษฐานะของครอบครัว แหล่งทรัพยากรการศึกษาที่บ้าน และความมั่งคั่งของครอบครัว ส่วนแบบสอบฉบับที่ 7 และ 8 ตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันจำนวน 4 ตัวแปร ได้แก่ เพศ เศรษฐานะของครอบครัว แหล่งทรัพยากรการศึกษาที่บ้าน และความมั่งคั่งของครอบครัว ส่วนแบบสอบฉบับที่ 13 ตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันจำนวน 4 ตัวแปร ได้แก่ การเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา เศรษฐานะของครอบครัว แหล่งทรัพยากรการศึกษาที่บ้าน และความมั่งคั่งของครอบครัว และแบบสอบฉบับที่ 4, 9, 10, และ 12 ตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันที่มีครบทั้ง 5 ตัวแปร ได้แก่ เพศ การเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา เศรษฐานะของครอบครัว แหล่งทรัพยากรการศึกษาที่บ้าน และความมั่งคั่งของครอบครัว รายละเอียดตามตารางที่ 36

ตารางที่ 36 ลักษณะข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันทซึ่งถูกตัดตามตัวแปรที่ศึกษา

แบบสอบ	ประเภท	เพศ	การเรียนพิเศษวิทยาศาสตร์	เศรษฐานะของครอบครัว	แหล่งทรัพยากรการศึกษาที่บ้าน	ความมั่งคั่งของครอบครัว	รวม
ฉบับที่ 2	Dicho	-	-	-	-	-	-
	Poly	-	-	✓	✓	✓	3
ฉบับที่ 3	Dicho	-	-	-	-	-	-
	Poly	-	-	✓	✓	✓	3
ฉบับที่ 4	Dicho	-	-	-	-	-	-
	Poly	✓	✓	✓	✓	✓	5
ฉบับที่ 7	Dicho	✓	-	-	-	-	1
	Poly	✓	-	✓	✓	✓	4
ฉบับที่ 8	Dicho	✓	-	-	-	-	1
	Poly	-	-	✓	✓	✓	3
ฉบับที่ 9	Dicho	✓	✓	-	-	-	2
	Poly	✓	✓	✓	✓	✓	5
ฉบับที่ 10	Dicho	✓	-	-	-	-	1
	Poly	-	✓	✓	✓	✓	4
ฉบับที่ 12	Dicho	✓	-	✓	✓	-	3
	Poly	✓	✓	✓	✓	✓	5
ฉบับที่ 13	Dicho	-	-	-	✓	-	1
	Poly	-	✓	✓	✓	✓	4

หมายเหตุ Dicho หมายถึง Dichotomous items, Poly หมายถึง Polytomous items

เมื่อตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันออกไปในแบบสอบแต่ละฉบับโดยยึดเกณฑ์การตัดข้อสอบไม่เกินร้อยละ 20 พบว่า แบบสอบฉบับที่ 4, 7, 9, 12, และ 13 มีร้อยละการตัดจำนวนข้อสอบมากที่สุด ร้อยละ 16.67 รองลงมาคือแบบสอบฉบับที่ 3 และ 10 ร้อยละ 14.29 และน้อยที่สุดคือ แบบสอบฉบับที่ 2 และ 8 ร้อยละ 11.76 และผลการตรวจสอบค่าความเที่ยงของข้อสอบก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า แบบสอบทุกฉบับมีค่าความเที่ยงตั้งแต่ .686 - .857 โดยมีแบบสอบฉบับที่ 12 มีค่าความเที่ยงมากที่สุด (.857) รองลงมาคือแบบสอบฉบับที่ 3 และ 10 ตามลำดับ (.844, .809) และแบบสอบฉบับที่ 4 มีค่าความเที่ยงน้อยที่สุด (.686) และค่าความเที่ยงของข้อสอบหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า แบบสอบทุกฉบับมีค่าความเที่ยงตั้งแต่ .620 - .821 โดยมีแบบสอบฉบับที่ 12 มีค่าความเที่ยงมากที่สุด (.821) รองลงมาคือ แบบสอบฉบับที่ 3 และ 10 ตามลำดับ (.809, .785) และ

แบบสอบฉบับที่ 4 มีค่าความเที่ยงน้อยที่สุด (.620) (ข้อสอบที่ถูกตัดทิ้งอยู่ในภาคผนวก ข) รายละเอียดตามตารางที่ 37

ตารางที่ 37 การเปรียบเทียบร้อยละข้อสอบที่ถูกตัด ค่าความเที่ยง และช่วงพารามิเตอร์ของข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันทั้ง 9 ฉบับ

ฉบับที่	จำนวนข้อสอบก่อนตัด			จำนวนข้อสอบที่ตัด					จำนวนข้อสอบ		ร้อยละข้อสอบที่เหลือ	ค่าความเที่ยง/(ช่วงพารามิเตอร์ความยาก)	
	ISI	EPS	USE	ISI	EPS	USE	รวม	ร้อยละ	ก่อนตัด	หลังตัด		ก่อนตัด	หลังตัด
2	6	5	6	-	-	2	2	11.76	17	15	88.24	.761 (-1.16-4.00)	.729 (-1.16-4.02)
3	9	11	15	-	1	4	5	14.29	35	30	85.71	.844 (-1.57-2.88)	.809 (-1.60-3.04)
4	4	11	3	1	1	1	3	16.67	18	15	83.33	.686 (-0.47-2.90)	.620 (-0.47-2.18)
7	3	6	9	-	1	2	3	16.67	18	15	83.33	.778 (-2.19-2.21)	.731 (-2.18-2.27)
8	6	5	6	-	-	2	2	11.76	17	15	88.24	.776 (-0.69-3.10)	.750 (-0.68-3.18)
9	4	11	3	1	1	1	3	16.67	18	15	83.33	.690 (-0.79-3.45)	.643 (-0.83-3.41)
10	10	16	9	1	1	3	5	14.29	35	30	85.71	.809 (-2.20-3.38)	.785 (-1.25-3.01)
12	7	17	12	1	2	3	6	16.67	36	30	83.33	.857 (-2.65-3.58)	.821 (-2.67-3.81)
13	3	6	9		1	2	3	16.67	18	15	83.33	.762 (-3.75-2.34)	.712 (-3.65-2.18)

หมายเหตุ สมรรถนะทางวิทยาศาสตร์ 3 ด้านคือ ISI หมายถึง การระบุประเด็นทางวิทยาศาสตร์, EPS หมายถึง การอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์, และ USE หมายถึง การใช้ประจักษ์พยานทางวิทยาศาสตร์

ผลการเปรียบเทียบร้อยละแบบสอบทำหน้าที่ต่างกันทั้งก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน โดยพิจาณาจากอิทธิพลขนาดกลาง (M) และขนาดใหญ่ (L) เป็นเกณฑ์ในการศึกษาการเกิดแบบสอบทำหน้าที่ต่างกัน พบว่า มีร้อยละแบบสอบที่ตรวจพบการทำหน้าที่ต่างกันลดลง 2 ฉบับ คือ ในแบบสอบฉบับที่ 12 ลดลงจากร้อยละ 40 เป็น 20 และแบบสอบฉบับที่ 13 ลดลงจากร้อยละ 20 เป็น 0 ส่วนอีก 7 ฉบับ พบว่า แบบสอบที่ตรวจพบการทำหน้าที่ต่างกันไม่เปลี่ยนแปลง โดยที่แบบสอบฉบับที่ 3, 7, และ 10 ยังคงมีร้อยละแบบสอบที่ตรวจพบการทำหน้าที่ต่างเท่าเดิม ส่วนแบบสอบฉบับที่ 2, 4, 8, และ 9 ไม่พบแบบสอบที่ตรวจพบการทำหน้าที่ต่างกัน และเมื่อพิจารณาตามตัวแปรที่ศึกษา พบว่า แบบสอบฉบับที่ 13 มี 2 ตัวแปร คือ แหล่งทรัพยากรการศึกษาที่บ้านและความมั่งคั่งของครอบครัวมีอิทธิพลแบบสอบที่ตรวจ

พบการทำหน้าที่ต่างกันลดลงจากขนาดกลางเป็นขนาดเล็ก และแบบสอบฉบับที่ 12 มี 1 ตัวแปร คือ ความมั่งคั่งของครอบครัวพบอิทธิพลแบบสอบที่ตรวจพบการทำหน้าที่ต่างกันลดลงจากขนาดกลางเป็นขนาดเล็ก รายละเอียดตามตารางที่ 38

ตารางที่ 38 ผลการตรวจสอบแบบสอบที่ตรวจพบการทำหน้าที่ต่างกันตามตัวแปรที่ศึกษา

ฉบับที่	รูปแบบการทำหน้าที่ต่างกันของแบบสอบตามตัวแปร										ร้อยละของ DTF (พิจารณา M-L)	
	ก่อนตัด					หลังตัด						
	GENDER	EXTRA	ESCS	HEDRES	WEALTH	GENDER	EXTRA	ESCS	HEDRES	WEALTH	ก่อนตัด	หลังตัด
2	S	S	S	S	S	S	S	S	S	S	0%	0%
3	S	S	M	M	M	S	S	M	M	M	60%	60%
4	S	S	S	S	S	S	S	S	S	S	0%	0%
7	S	S	M	S	S	S	S	M	S	S	20%	20%
8	S	S	S	S	S	S	S	S	S	S	0%	0%
9	S	S	S	S	S	S	S	S	S	S	0%	0%
10	S	S	M	S	S	S	S	M	S	S	20%	20%
12	S	S	M	S	M	S	S	M	S	S	40%	20%
13	S	S	S	M	M	S	S	S	S	S	40%	0%

ภายหลังผู้วิจัยได้ตรวจสอบความตรงเชิงโครงสร้างด้วยการวิเคราะห์องค์ประกอบเชิงยืนยัน เพื่อทดสอบความสอดคล้องของโมเดลตามสมรรถนะทางวิทยาศาสตร์ โดยการยึดองค์ประกอบตามสมรรถนะทางวิทยาศาสตร์ 3 ด้าน ได้แก่ การระบุประเด็นทางวิทยาศาสตร์ (ISI) การอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ (EPS) และการใช้ประจักษ์พยานทางวิทยาศาสตร์ (USE) ก่อนที่จะทำการวิเคราะห์องค์ประกอบเชิงยืนยัน ผู้วิจัยได้ศึกษาความสัมพันธ์ระหว่างองค์ประกอบทั้ง 3 ด้านในแบบสอบแต่ละฉบับและภาพรวมทุกฉบับ เพื่อพิจารณาความเหมาะสมของเมทริกสหสัมพันธ์ที่จะนำไปสู่การวิเคราะห์องค์ประกอบ โดยใช้สถิติของ Bartlett เป็นสถิติที่ใช้ทดสอบสมมติฐานความเป็นเมทริกเอกลักษณ์ (identity matrix) และค่าดัชนีไกเซอร์-ไมเยอร์-ออลคิน (Kaiser-Meyer-Olkin Measure of Sampling Adequacy=MSA) ทดสอบความเหมาะสมของข้อมูลทั้งสองช่วง คือ ก่อนและหลังตัดข้อสอบที่ตรวจพบการหน้าที่ต่างกันตามเกณฑ์ที่กำหนดไว้ ทั้งนี้เพื่อเป็นการยืนยันความตรงเชิงโครงสร้างของข้อสอบทั้งสองช่วงการวิจัยซึ่งยังคงมีโครงสร้างข้อสอบที่ครบถ้วนตามคุณสมบัติของการเป็นข้อสอบที่ดี

ผลการพิจารณาภาพรวมของแบบสอบก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า มีค่า Bartlett's test of sphericity เท่ากับ 3182.143 (p<.000) และค่าดัชนี ไกเซอร์-ไมเยอร์-ออลคิน เท่ากับ .524 ส่วนหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า มีค่า Bartlett's test of sphericity เท่ากับ 2226.205 (p<.000) และค่าดัชนีไกเซอร์-ไมเยอร์-ออลคิน เท่ากับ .566 ส่วนความสัมพันธ์ระหว่างสมรรถนะทางวิทยาศาสตร์ พบว่า ทั้งก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ซึ่งก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันระหว่างสมรรถนะการระบุประเด็นทางวิทยาศาสตร์ (ISI) กับการอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ (EPS) มีความสัมพันธ์สูงสุด (.571) ส่วนหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันระหว่างสมรรถนะการระบุประเด็นทางวิทยาศาสตร์ (ISI) กับการใช้ประจักษ์พยานทางวิทยาศาสตร์ (USE) มีความสัมพันธ์สูงสุด (.501)

ส่วนแบบสอบฉบับที่ 2 ก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า มีค่า Bartlett's test of sphericity เท่ากับ 392.425 (p<.000) และค่าดัชนีไกเซอร์-ไมเยอร์-ออลคิน เท่ากับ .684 ส่วนหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า มีค่า Bartlett's test of sphericity เท่ากับ 392.614 (p<.000) และค่าดัชนีไกเซอร์-ไมเยอร์-ออลคิน เท่ากับ .683 ส่วนความสัมพันธ์ระหว่างสมรรถนะทางวิทยาศาสตร์ พบว่า ทั้งก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ซึ่งก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันระหว่างสมรรถนะการระบุประเด็นทางวิทยาศาสตร์ (ISI) กับการใช้ประจักษ์พยานทางวิทยาศาสตร์ (USE) มีความสัมพันธ์สูงสุด (.585) ส่วนหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันกันระหว่างสมรรถนะการระบุประเด็นทางวิทยาศาสตร์ (ISI) กับการอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ (EPS) มีความสัมพันธ์สูงสุด (.585)

แบบสอบฉบับที่ 3 ก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า มีค่า Bartlett's test of sphericity เท่ากับ 490.009 (p<.000) และค่าดัชนีไกเซอร์-ไมเยอร์-ออลคิน เท่ากับ .686 ส่วนหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า มีค่า Bartlett's test of sphericity เท่ากับ 453.415 (p<.000) และค่าดัชนีไกเซอร์-ไมเยอร์-ออลคิน เท่ากับ .679 ส่วนความสัมพันธ์ระหว่างสมรรถนะทางวิทยาศาสตร์ พบว่า ทั้งก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ซึ่งก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันระหว่างสมรรถนะการอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ (EPS) กับการใช้ประจักษ์พยานทางวิทยาศาสตร์ (USE) มีความสัมพันธ์สูงสุด (.673) ส่วนหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันระหว่างสมรรถนะการอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ (EPS) กับการใช้ประจักษ์พยานทางวิทยาศาสตร์ (USE) มีความสัมพันธ์สูงสุดเช่นเดียวกัน (.659)

แบบสอบฉบับที่ 4 ก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า มีค่า Bartlett's test of sphericity เท่ากับ 232.381 (p<.000) และค่าดัชนีไกเซอร์-ไมเยอร์-ออลคิน เท่ากับ .644 ส่วนหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า มีค่า Bartlett's test of sphericity เท่ากับ 1167.260 (p<.000) และค่าดัชนีไกเซอร์-ไมเยอร์-ออลคิน เท่ากับ .756 ส่วนความสัมพันธ์ระหว่างสมรรถนะทางวิทยาศาสตร์ พบว่า ทั้งก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ซึ่งก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันระหว่างสมรรถนะการอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ (EPS) กับการใช้ประจักษ์พยานทางวิทยาศาสตร์ (USE) มีความสัมพันธ์สูงสุด (.487) ส่วนหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันระหว่างสมรรถนะการระบุประเด็นทางวิทยาศาสตร์ (ISI) กับการใช้ประจักษ์พยานทางวิทยาศาสตร์ (USE) มีความสัมพันธ์สูงสุด (.861)

แบบสอบฉบับที่ 7 ก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า มีค่า Bartlett's test of sphericity เท่ากับ 284.608 (p<.000) และค่าดัชนีไกเซอร์-ไมเยอร์-ออลคิน เท่ากับ .623 ส่วนหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า มีค่า Bartlett's test of sphericity เท่ากับ 255.162 (p<.000) และค่าดัชนีไกเซอร์-ไมเยอร์-ออลคิน เท่ากับ .627 ส่วนความสัมพันธ์ระหว่างสมรรถนะทางวิทยาศาสตร์ พบว่า ทั้งก่อนและหลังตัดข้อสอบที่ตรวจพบการหน้าที่ต่างกันมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ซึ่งก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันระหว่างสมรรถนะการอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ (EPS) กับการใช้ประจักษ์พยานทางวิทยาศาสตร์ (USE) มีความสัมพันธ์สูงสุด (.579) ส่วนหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันระหว่างสมรรถนะการอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ (EPS) กับการใช้ประจักษ์พยานทางวิทยาศาสตร์ (USE) มีความสัมพันธ์สูงสุดเช่นเดียวกัน (.532)

แบบสอบฉบับที่ 8 ก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า มีค่า Bartlett's test of sphericity เท่ากับ 350.408 (p<.000) และค่าดัชนีไกเซอร์-ไมเยอร์-ออลคิน เท่ากับ .675 ส่วนหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า มีค่า Bartlett's test of sphericity เท่ากับ 317.551 (p<.000) และค่าดัชนีไกเซอร์-ไมเยอร์-ออลคิน เท่ากับ .677 ส่วนความสัมพันธ์ระหว่างสมรรถนะทางวิทยาศาสตร์ พบว่า ทั้งก่อนและหลังตัดข้อสอบที่ตรวจพบการหน้าที่ต่างกันมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ซึ่งก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันระหว่างสมรรถนะการระบุประเด็นทางวิทยาศาสตร์ (ISI) กับการใช้ประจักษ์พยานทางวิทยาศาสตร์ (USE) มีความสัมพันธ์สูงสุด (.557) ส่วนหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันระหว่างสมรรถนะการระบุประเด็นทางวิทยาศาสตร์ (ISI) กับการใช้ประจักษ์พยานทางวิทยาศาสตร์ (USE) มีความสัมพันธ์สูงสุดเช่นเดียวกัน (.530)

แบบสอบฉบับที่ 9 ก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า มีค่า Bartlett's test of sphericity เท่ากับ 195.414 (p<.000) และค่าดัชนีไกเซอร์-ไมเยอร์-ออลคิน เท่ากับ .637 ส่วนหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า มีค่า Bartlett's test of sphericity เท่ากับ 124.415 (p<.000) และค่าดัชนีไกเซอร์-ไมเยอร์-ออลคิน เท่ากับ .602 ส่วนความสัมพันธ์ระหว่างสมรรถนะทางวิทยาศาสตร์ พบว่า ทั้งก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ซึ่งก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันระหว่างสมรรถนะการอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ (EPS) กับการใช้ประจักษ์พยานทางวิทยาศาสตร์ (USE) มีความสัมพันธ์สูงสุด (.439) ส่วนหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันระหว่างสมรรถนะการระบุประเด็นทางวิทยาศาสตร์ (ISI) กับการอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ (EPS) มีความสัมพันธ์สูงสุด (.354)

แบบสอบฉบับที่ 10 ก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า มีค่า Bartlett's test of sphericity เท่ากับ 425.688 (p<.000) และค่าดัชนีไกเซอร์-ไมเยอร์-ออลคิน เท่ากับ .703 ส่วนหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า มีค่า Bartlett's test of sphericity เท่ากับ 367.429 (p<.000) และค่าดัชนีไกเซอร์-ไมเยอร์-ออลคิน เท่ากับ .699 ส่วนความสัมพันธ์ระหว่างสมรรถนะทางวิทยาศาสตร์ พบว่า ทั้งก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ซึ่งก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันระหว่างสมรรถนะการอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ (EPS) กับการใช้ประจักษ์พยานทางวิทยาศาสตร์ (USE) มีความสัมพันธ์สูงสุด (.601) ส่วนหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันระหว่างสมรรถนะการระบุประเด็นทางวิทยาศาสตร์ (ISI) กับการใช้ประจักษ์พยานทางวิทยาศาสตร์ (USE) มีความสัมพันธ์สูงสุด (.548)

แบบสอบฉบับที่ 12 ก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า มีค่า Bartlett's test of sphericity เท่ากับ 560.992 (p<.000) และค่าดัชนีไกเซอร์-ไมเยอร์-ออลคิน เท่ากับ .709 ส่วนหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า มีค่า Bartlett's test of sphericity เท่ากับ 474.550 (p<.000) และค่าดัชนีไกเซอร์-ไมเยอร์-ออลคิน เท่ากับ .708 ส่วนความสัมพันธ์ระหว่างสมรรถนะทางวิทยาศาสตร์ พบว่า ทั้งก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าหน้าที่ต่างกันมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ซึ่งก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันระหว่างสมรรถนะการอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ (EPS) กับการใช้ประจักษ์พยานทางวิทยาศาสตร์ (USE) มีความสัมพันธ์สูงสุด (.698) ส่วนหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันระหว่างสมรรถนะการอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ (EPS) กับการใช้ประจักษ์พยานทางวิทยาศาสตร์ (USE) มีความสัมพันธ์สูงสุดเช่นเดียวกัน (.626)

และแบบสอบฉบับที่ 13 ก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า มีค่า Bartlett's test of sphericity เท่ากับ 278.768 (p<.000) และค่าดัชนีไกเซอร์-ไมเยอร์-ออลคิน เท่ากับ .638 ส่วนหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า มีค่า Bartlett's test of sphericity เท่ากับ 236.400 (p<.000) และค่าดัชนีไกเซอร์-ไมเยอร์-ออลคิน เท่ากับ .638 ส่วนความสัมพันธ์ระหว่างสมรรถนะทางวิทยาศาสตร์ พบว่า ทั้งก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าหน้าที่ต่างกันมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ซึ่งก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันระหว่างสมรรถนะการอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ (EPS) กับการใช้ประจักษ์พยานทางวิทยาศาสตร์ (USE) มีความสัมพันธ์สูงสุด (.557) ส่วนหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันระหว่างสมรรถนะการอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ (EPS) กับการใช้ประจักษ์พยานทางวิทยาศาสตร์ (USE) มีความสัมพันธ์สูงสุดเช่นเดียวกัน (.514)

เมื่อพิจารณาแบบสอบภาพรวมทุกฉบับและรายฉบับก่อนตัดข้อสอบที่ตรวจพบการทำหน้าหน้าที่ต่างกัน พบว่า เมทริกสหสัมพันธ์ระหว่างตัวบ่งชี้แตกต่างจากเมทริกเอกลักษณ์อย่างมีนัยสำคัญทางสถิติและค่าดัชนีไกเซอร์-ไมเยอร์-ออลคิน(Kaiser-Meyer-Olkin Measure of Sampling Adequacy) มีค่าตั้งแต่ .566 -.709 ซึ่งมากกว่า .5 มีความเหมาะสมที่จะนำมาวิเคราะห์องค์ประกอบ ส่วนแบบสอบภาพรวมทุกฉบับและรายฉบับหลังตัดข้อสอบที่ตรวจพบการทำหน้าหน้าที่ต่างกัน พบว่า เมทริกสหสัมพันธ์ระหว่างตัวบ่งชี้แตกต่างจากเมทริกเอกลักษณ์อย่างมีนัยสำคัญทางสถิติ และค่าดัชนีไกเซอร์-ไมเยอร์-ออลคิน (Kaiser-Meyer-Olkin Measure of Sampling Adequacy) มีค่าตั้งแต่ .524 - .708 มากกว่า .5 หมายความว่า แบบสอบภาพรวมทุกฉบับและรายฉบับหลังตัดข้อสอบที่ตรวจพบการทำหน้าหน้าที่ต่างกันไปแล้วยังคงมีความตรงเชิงโครงสร้างตามองค์ประกอบสมรรถนะทางวิทยาศาสตร์ รายละเอียดตามตารางที่ 39

ตารางที่ 39 สัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรตามสมรรถนะทางวิทยาศาสตร์

แบบสอบ	สมรรถนะ		Mean	SD	ความสัมพันธ์			Bartlett's Test of Sphericity	df	p	MSA
					ISI–EPS	ISI–USE	EPS-USE				
2	ก่อนตัดข้อสอบ DIF	ISI	.414	.242	.457**	.585**	.547**	392.425**	3	.000	.684
		EPS	.392	.246							
		USE	.471	.285							
	หลังตัดข้อสอบ DIF	ISI	.414	.242	.585**	.457**	.548**	392.614**	3	.000	.683
		EPS	.392	.246							
		USE	.471	.285							
3	ก่อนตัดข้อสอบ DIF	ISI	.411	.199	.496**	.576**	.673**	490.009**	3	.000	.686
		EPS	.382	.197							
		USE	.425	.245							
	หลังตัดข้อสอบ DIF	ISI	.411	.199	.474**	.548**	.659**	453.415**	3	.000	.679

ตารางที่ 39 สัมประสิทธิ์สหสัมพันธ์ของตัวแปรตามสมรรถนะทางวิทยาศาสตร์ (ต่อ)

แบบสอบ	สมรรถนะ		Mean	SD	ความสัมพันธ์			Bartlett's Test of Sphericity	df	p	MSA
4	ก่อนตัดข้อสอบ DIF	ISI	.350	.259	.425**	.330**	.487**	232.381**	3	.000	.644
		EPS	.409	.194							
		USE	.368	.344							
	หลังตัดข้อสอบ DIF	ISI	.375	.663	.800**	.861**	.786**	1167.260**	3	.000	.756
		EPS	.482	.355							
		USE	.417	1.167							
7	ก่อนตัดข้อสอบ DIF	ISI	.347	.308	.319**	.407**	.579**	284.608**	3	.000	.623
		EPS	.401	.225							
		USE	.457	.294							
	หลังตัดข้อสอบ DIF	ISI	.347	.308	.309**	.420**	.532**	255.162**	3	.000	.627
		EPS	.448	.241							
		USE	.351	.254							
8	ก่อนตัดข้อสอบ DIF	ISI	.385	.242	.426**	.557**	.534**	350.408**	3	.000	.675
		EPS	.340	.259							
		USE	.417	.588							
	หลังตัดข้อสอบ DIF	ISI	.385	.242	.426**	.530**	.505**	317.551**	3	.000	.677
		EPS	.340	.259							
		USE	.442	.301							
9	ก่อนตัดข้อสอบ DIF	ISI	.399	.256	.403**	.304**	.439**	195.414**	3	.000	.637
		EPS	.430	.204							
		USE	.437	.370							
	หลังตัดข้อสอบ DIF	ISI	.319	.274	.354**	.212**	.212**	124.415**	3	.000	.602
		EPS	.401	.216							
		USE	.294	.337							
10	ก่อนตัดข้อสอบ DIF	ISI	.392	.196	.521**	.563**	.601**	425.688**	3	.000	.703
		EPS	.402	.181							
		USE	.411	.234							
	หลังตัดข้อสอบ DIF	ISI	.337	.213	.523**	.548**	.516**	367.429**	3	.000	.699
		EPS	.414	.187							
		USE	.347	.235							
12	ก่อนตัดข้อสอบ DIF	ISI	.380	.226	.594**	.571**	.698**	560.992**	3	.000	.709
		EPS	.398	.179							
		USE	.467	.269							
	หลังตัดข้อสอบ DIF	ISI	.347	.237	.586**	.542**	.626**	474.550**	3	.000	.708
		EPS	.410	.198							
		USE	.354	.229							

ตารางที่ 39 สัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรตามสมรรถนะทางวิทยาศาสตร์ (ต่อ)

แบบสอบ	สมรรถนะ		Mean	SD	ความสัมพันธ์			Bartlett's Test of Sphericity	df	p	MSA
					ISI–EPS	ISI–USE	EPS-USE				
13	ก่อนตัดข้อสอบ DIF	ISI	.405	.300	.342**	.407**	.557**	278.768**	3	.000	.638
		EPS	.419	.219							
		USE	.507	.294							
	หลังตัดข้อสอบ DIF	ISI	.405	.300	.336**	.371**	.514**	236.400**	3	.000	.638
		EPS	.460	.230							
		USE	.398	.266							
รวมทุกฉบับ	ก่อนตัดข้อสอบ DIF	ISI	4.148	1.208	.571**	.529**	.207**	3182.143**	3	0.00	.524
		EPS	3.932	1.290							
		USE	4.140	1.346							
	หลังตัดข้อสอบ DIF	ISI	5.259	1.666	.453**	.501**	.212**	2226.205**	3	.000	.566
		EPS	5.262	1.660							
		USE	5.234	1.896							

หมายเหตุ ** หมายถึง p<.01

ส่วนผลการตรวจสอบความตรงตามทฤษฎีหรือความสอดคล้องของโมเดลสมรรถนะทางวิทยาศาสตร์โดยใช้การวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า โมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ เมื่อพิจารณาจากค่าไค-สแควร์ (Chi-square) ของแบบสอบภาพรวมทุกฉบับ เท่ากับ 1.980 ที่องศาความเป็นอิสระ 1.980 ค่าดัชนีวัดระดับความกลมกลืน (GFI) เท่ากับ 1.000 ค่าดัชนีวัดระดับความกลมกลืนที่ปรับแก้แล้ว (AGFI) เท่ากับ 1.000 และรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของส่วนเหลือ (RMR) เท่ากับ 0.012 นั่นคือ ยอมรับสมมติฐานหลักที่ว่าโมเดลแบบสอบภาพรวมทุกฉบับก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันมีความสอดคล้องกับข้อมูลเชิงประจักษ์ และเมื่อพิจารณาแยกแบบสอบทั้ง 9 ฉบับ พบว่า จากค่าไค-สแควร์ (Chi-square) มีค่าอยู่ระหว่าง 1.220 – 1.910 องศาความเป็นอิสระที่ 1.000 โดยค่าดัชนีวัดระดับความกลมกลืน (GFI) ทั้งหมดมีค่า 1.000 ทั้งหมด ค่าดัชนีวัดระดับความกลมกลืนที่ปรับแก้แล้ว (AGFI) ทั้งหมดอยู่ระหว่าง 0.980 - 0.990 และรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของส่วนเหลือ (RMR) อยู่ระหว่าง 0.001 – 0.002 นั่นคือ ยอมรับสมมติฐานหลักที่ว่าโมเดลของแบบสอบทั้ง 9 ฉบับก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ทั้ง 9 ฉบับ รายละเอียดตามตารางที่ 40

ตารางที่ 40 สรุปดัชนีที่ใช้ในการตรวจสอบความสอดคล้องกลมกลืนของโมเดลกับข้อมูลเชิงประจักษ์โมเดลก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน

ดัชนีทดสอบ	แบบสอบ									
	ฉบับที่ 2	ฉบับที่ 3	ฉบับที่ 4	ฉบับที่ 7	ฉบับที่ 8	ฉบับที่ 9	ฉบับที่10	ฉบับที่12	ฉบับที่13	รวม
n	484	475	473	476	475	473	471	478	487	**4,292**
$\chi 2$	1.770	1.310	1.670	1.730	1.910	1.220	1.740	1.520	1.410	**1.980**
df.	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	**1.000**
P-value	0.183	0.252	0.197	0.188	0.167	0.270	0.187	0.218	0.235	**0.160**
$\chi 2/df$	1.770	1.310	1.670	1.730	1.910	1.220	1.740	1.520	1.410	**1.980**
CFI	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	**1.000**
GFI	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	**1.000**
AGFI	0.980	0.990	0.990	0.990	0.980	0.990	0.990	0.990	0.990	**1.000**
RMR	0.001	0.017	0.001	0.002	0.001	0.001	0.001	0.001	0.002	**0.012**
สรุป	สอดคล้อง	สอดคล้อง	สอดคล้อง	สอดคล้อง	สอดคล้อง	สอดคล้อง	สอดคล้อง	สอดคล้อง	สอดคล้อง	สอดคล้อง

หมายเหตุ ดัชนีและเกณฑ์ที่ใช้ตัดสินความสอดคล้องกลมกลืนของโมเดลกับข้อมูลเชิงประจักษ์

$\chi 2$ = Chi-Square Statistics (ไม่มีนัยสำคัญ, $p<.05$)

$\chi 2/df$ = Chi-Square Statistics / degree of freedom ($\chi 2/df < 2.0$)

CFI = Comparative Fit Index (CFI ≥ 0.95)

GFI = Goodness of Fit Index (GFI ≥ 0.95)

AGFI = Adjusted Goodness of Fit Index (AGFI ≥ 0.95)

RMR = Root Mean Square Residual (RMR < .05)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ส่วนผลการตรวจสอบความตรงตามทฤษฎีหรือความสอดคล้องของโมเดลสมรรถนะทางวิทยาศาสตร์โดยใช้การวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลภายหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า โมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ เมื่อพิจารณาจากค่าไค-สแควร์ (Chi-square) ของแบบสอบภาพรวมทุกฉบับ เท่ากับ 1.980 ที่องศาความเป็นอิสระ 1.600 ค่าดัชนีวัดระดับความกลมกลืน (GFI) เท่ากับ 1.000 ค่าดัชนีวัดระดับความกลมกลืนที่ปรับแก้แล้ว (AGFI) เท่ากับ 1.000 และรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของส่วนเหลือ (RMR) เท่ากับ 0.019 นั่นคือ ยอมรับสมมติฐานหลักที่ว่าโมเดลแบบสอบภาพรวมทุกฉบับภายหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันมีความสอดคล้องกับข้อมูลเชิงประจักษ์ และเมื่อพิจารณาแยกแบบสอบทั้ง 9 ฉบับ พบว่า จากค่าไค-สแควร์ (Chi-square) มีค่าอยู่ระหว่าง 1.190 – 1.980 องศาความเป็นอิสระที่ 1.000 ทั้งนี้ค่าดัชนีวัดระดับความกลมกลืน (GFI) ทั้งหมดมีค่า 1.000 ทั้งหมด ค่าดัชนีวัดระดับความกลมกลืนที่ปรับแก้แล้ว (AGFI) ทั้งหมดอยู่ระหว่าง 0.950 -

0.990 และรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของส่วนเหลือ (RMR) อยู่ระหว่าง 0.001 – 0.023 นั่นคือ ยอมรับสมมติฐานหลักที่ว่าโมเดลของแบบสอบทั้ง 9 ฉบับภายหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ทั้ง 9 ฉบับ รายละเอียดตามตารางที่ 41

ตารางที่ 41 สรุปดัชนีที่ใช้ในการตรวจสอบความสอดคล้องกลมกลืนของโมเดลกับข้อมูลเชิงประจักษ์ของโมเดลหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน

ดัชนีทดสอบ	แบบสอบ									
	ฉบับที่ 2	ฉบับที่ 3	ฉบับที่ 4	ฉบับที่ 7	ฉบับที่ 8	ฉบับที่ 9	ฉบับที่10	ฉบับที่12	ฉบับที่13	รวม
n	484	475	473	476	475	473	471	478	487	**4,292**
$\chi 2$	1.980	1.290	1.380	1.770	1.670	1.190	1.900	1.740	1.340	**1.600**
df.	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	**1.000**
P-value	0.160	0.225	0.240	0.184	0.184	0.276	0.163	0.187	0.248	**0.206**
$\chi 2/df$	1.980	1.290	1.380	1.770	1.670	1.190	1.900	1.740	1.340	**1.600**
CFI	1.000	0.990	1.000	1.000	1.000	1.000	1.000	1.000	1.000	**1.000**
GFI	1.000	0.990	1.000	1.000	1.000	1.000	1.000	1.000	1.000	**1.000**
AGFI	0.980	0.950	0.990	0.990	0.990	0.990	0.980	0.980	0.990	**1.000**
RMR	0.001	0.001	0.002	0.023	0.001	0.002	0.001	0.001	0.001	**0.019**
สรุป	สอดคล้อง	สอดคล้อง	สอดคล้อง	สอดคล้อง	สอดคล้อง	สอดคล้อง	สอดคล้อง	สอดคล้อง	สอดคล้อง	สอดคล้อง

หมายเหตุ ดัชนีและเกณฑ์ที่ใช้ตัดสินความสอดคล้องกลมกลืนของโมเดลกับข้อมูลเชิงประจักษ์

$\chi 2$ = Chi-Square Statistics (ไม่มีนัยสำคัญ, p<.05)

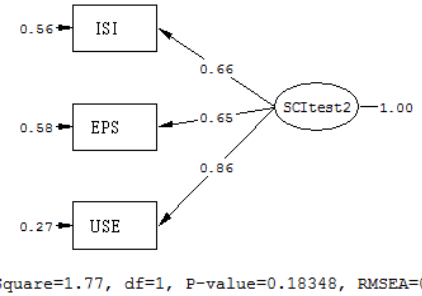
$\chi 2/df$ = Chi-Square Statistics / degree of freedom ($\chi 2/df$ < 2.0)

CFI = Comparative Fit Index (CFI $\geq$ 0.95)

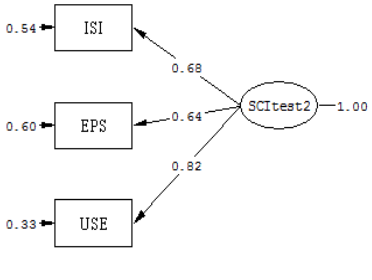
GFI = Goodness of Fit Index (GFI $\geq$ 0.95)

AGFI = Adjusted Goodness of Fit Index (AGFI $\geq$ 0.95)

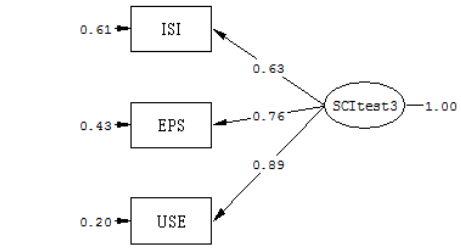
RMR = Root Mean Square Residual (RMR < .05)



โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันโมเดลก่อนตัดข้อสอบ DIF ของแบบสอบที่ 2

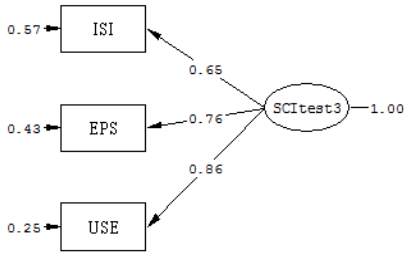


โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันโมเดลหลังตัดข้อสอบ DIF ของแบบสอบที่ 2



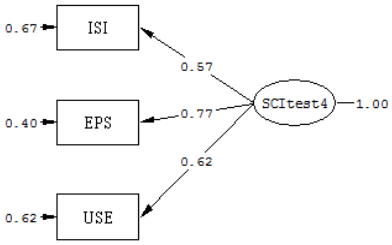
Chi-Square=1.31, df=1, P-value=0.25156, RMSEA=0.026

โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันโมเดล
ก่อนตัดข้อสอบ DIF ของแบบสอบที่ 3



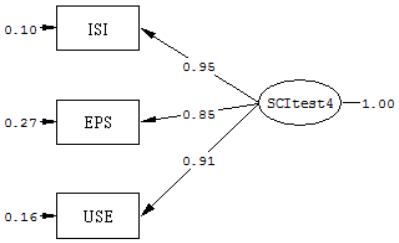
Chi-Square=1.29, df=1, P-value=0.25530, RMSEA=0.025

โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันโมเดล
หลังตัดข้อสอบ DIF ของแบบสอบที่ 3



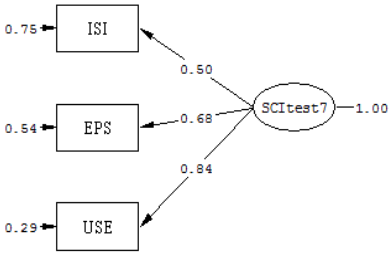
Chi-Square=1.67, df=1, P-value=0.19665, RMSEA=0.038

โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันโมเดล
ก่อนตัดข้อสอบ DIF ของแบบสอบที่ 4



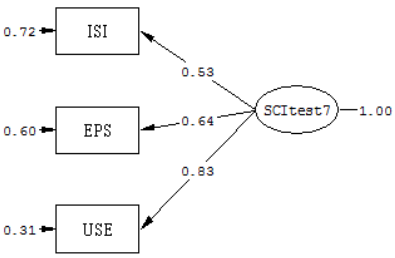
Chi-Square=1.38, df=1, P-value=0.23994, RMSEA=0.028

โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันโมเดล
หลังตัดข้อสอบ DIF ของแบบสอบที่ 4



Chi-Square=1.73, df=1, P-value=0.18793, RMSEA=0.039

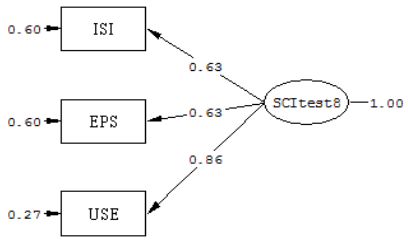
โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันโมเดล
ก่อนตัดข้อสอบ DIF ของแบบสอบที่ 7



Chi-Square=1.77, df=1, P-value=0.18376, RMSEA=0.040

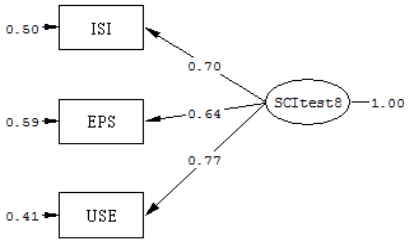
โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันโมเดล
หลังตัดข้อสอบ DIF ของแบบสอบที่ 7

Chulalongkorn University



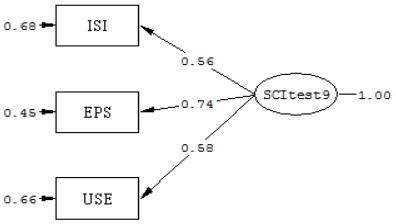
Chi-Square=1.91, df=1, P-value=0.16706, RMSEA=0.044

โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันโมเดล
ก่อนตัดข้อสอบ DIF ของแบบสอบที่ 8



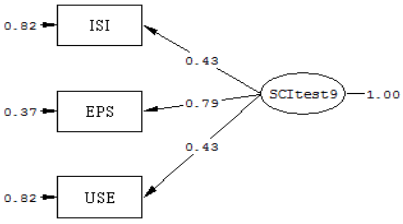
Chi-Square=1.67, df=1, P-value=0.19689, RMSEA=0.037

โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันโมเดล
หลังตัดข้อสอบ DIF ของแบบสอบที่ 8



Chi-Square=1.22, df=1, P-value=0.26958, RMSEA=0.022

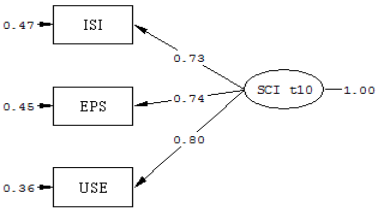
โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันโมเดล
ก่อนตัดข้อสอบ DIF ของแบบสอบที่ 9



Chi-Square=1.19, df=1, P-value=0.27562, RMSEA=0.020

โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันโมเดล
หลังตัดข้อสอบ DIF ของแบบสอบที่ 9

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

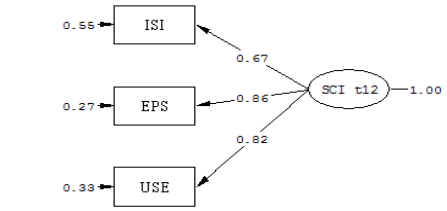


Chi-Square=1.74, df=1, P-value=0.18686, RMSEA=0.040

โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันโมเดล
ก่อนตัดข้อสอบ DIF ของแบบสอบที่ 10

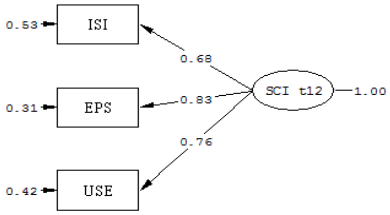
Chi-Square=1.90, df=1, P-value=0.16839, RMSEA=0.044

โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันโมเดล
หลังตัดข้อสอบ DIF ของแบบสอบที่ 10



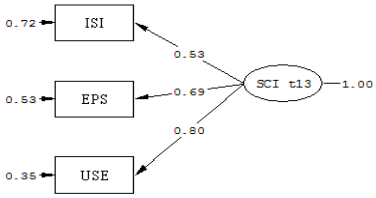
Chi-Square=1.52, df=1, P-value=0.21787, RMSEA=0.033

โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันโมเดล
ก่อนตัดข้อสอบ DIF ของแบบสอบที่ 12



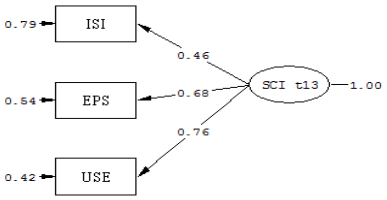
Chi-Square=1.74, df=1, P-value=0.18690, RMSEA=0.039

โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันโมเดล
หลังตัดข้อสอบ DIF ของแบบสอบที่ 12



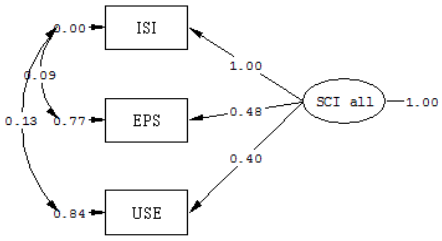
Chi-Square=1.41, df=1, P-value=0.23452, RMSEA=0.029

โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันโมเดล
ก่อนตัดข้อสอบ DIF ของแบบสอบที่ 13



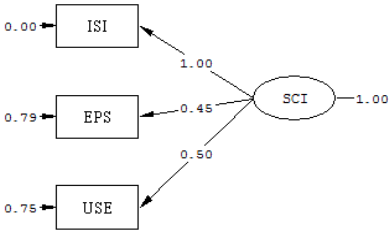
Chi-Square=1.34, df=1, P-value=0.24758, RMSEA=0.026

โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันโมเดล
หลังตัดข้อสอบ DIF ของแบบสอบที่ 13



Chi-Square=1.98, df=1, P-value=0.15978, RMSEA=0.015

โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันโมเดล
ก่อนตัดข้อสอบ DIF ของแบบสอบรวมทุกฉบับ



Chi-Square=1.60, df=1, P-value=0.20575, RMSEA=0.012

โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันโมเดล
หลังตัดข้อสอบ DIF ของแบบสอบรวมทุกฉบับ

รูปภาพที่ 12 แสดงโมเดลวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลก่อนและหลังตัดข้อสอบ DIF ในแบบสอบ 9 ฉบับและรวมทุกฉบับ

2.3 ผลการประมาณค่าคะแนนความสามารถของนักเรียนหลังจากตรวจสอบโครงสร้างข้อสอบ

หลังการตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันออกไปตามเกณฑ์ที่กำหนด ซึ่งผู้วิจัยดำเนินการประมาณค่าคะแนนความสามารถของนักเรียนเป็นจำนวน 2 ครั้ง คือ ครั้งแรกเป็นการประมาณค่าคะแนนความสามารถของนักเรียนก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน และครั้งที่สองเป็นการประมาณค่าคะแนนความสามารถของนักเรียนหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน โดยทั้งสองครั้งมีการแปลงค่าคะแนนความสามารถของนักเรียนเป็นคะแนนตามสเกลของ PISA (รายละเอียดในบทที่ 3)

ผลการประมาณค่าคะแนนความสามารถของนักเรียนก่อนการตัดข้อสอบที่ตรวจพบทำหน้าที่ต่างกัน พบว่า ภาพรวมทั้ง 9 ฉบับ มีค่าความสามารถโดยเฉลี่ยเท่ากับ 0.001 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.879 โดยมีนักเรียนที่เข้าสอบได้ค่าความสามารถต่ำสุดเท่ากับ -2.206 และสูงสุดเท่ากับ 3.034 โดยภาพรวมคะแนนความสามารถของนักเรียนมีคะแนนความสามารถต่ำกว่าความสามารถเฉลี่ยและต่ำกว่าโค้งปกติ (Sk=0.303, Ku=-.218) เมื่อพิจารณาคะแนนความสามารถของนักเรียนที่ทำแบบสอบในแต่ละฉบับ พบว่า แบบสอบฉบับที่ 4 มีคะแนนความสามารถเฉลี่ยของนักเรียนสูงสุด (Mean = 0.003, SD = 0.832) และแบบสอบฉบับที่ 3 มีคะแนนความสามารถเฉลี่ยของนักเรียนต่ำสุด (Mean = -0.002, SD = 0.915) รายละเอียดตามตารางที่ 42

ตารางที่ 42 ค่าสถิติพื้นฐานคะแนนความสามารถของนักเรียนก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันในแบบสอบ 9 ฉบับ

ฉบับที่	N	Range	Min	Max	Mean	SD	Var	Sk	Ku
2	484	3.937	-1.882	2.055	-0.001	0.854	0.730	0.273	-0.349
3	475	4.793	-2.008	2.785	-0.002	0.915	0.837	0.624	0.185
4	473	4.514	-1.866	2.648	0.003	0.832	0.692	0.230	0.045
7	476	4.506	-1.967	2.539	0.002	0.881	0.776	0.235	-0.379
8	475	4.237	-1.761	2.476	0.000	0.877	0.769	0.148	-0.415
9	473	4.492	-1.999	2.493	0.001	0.834	0.695	0.237	-0.221
10	471	4.766	-2.029	2.737	-0.001	0.900	0.811	0.391	-0.070
12	478	5.240	-2.206	3.034	0.002	0.928	0.860	0.338	-0.256
13	487	4.398	-2.008	2.390	0.002	0.874	0.764	0.173	-0.584
รวม	**4,292**	**5.240**	**-2.206**	**3.034**	**0.001**	**0.877**	**0.769**	**0.302**	**-0.219**

ส่วนผลการประมาณค่าคะแนนความสามารถของนักเรียนหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันออกไป พบว่า ภาพรวมทั้ง 9 ฉบับ มีคะแนนความสามารถเฉลี่ยเท่ากับ -0.001 ส่วนเบี่ยงเบนมาตรฐาน 0.856 โดยที่นักเรียนที่เข้าสอบมีคะแนนความสามารถต่ำสุดเท่ากับ -2.337 และสูงสุดเท่ากับ 2.756 โดยภาพรวมคะแนนความสามารถของนักเรียนมีคะแนนความสามารถต่ำกว่าค่าความสามารถเฉลี่ยและต่ำกว่าโค้งปกติ (Sk=0.325, Ku=-0.216) เมื่อพิจารณาคะแนนความสามารถของนักเรียนที่ทำแบบสอบแต่ละฉบับ พบว่า แบบสอบฉบับที่ 4 มีคะแนนความสามารถเฉลี่ยของนักเรียนสูงสุด (Mean = 0.001, SD = 0.791) และแบบสอบฉบับที่ 10 มีคะแนนความสามารถเฉลี่ยของนักเรียนต่ำสุด (Mean = -0.003, SD = 0.889) รายละเอียดตามตารางที่ 43

ตารางที่ 43 ค่าสถิติพื้นฐานคะแนนความสามารถของนักเรียนหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันในแบบสอบ 9 ฉบับ

ฉบับที่	N	Range	Min	Max	Mean	SD	Var	Sk	Ku
2	484	4.290	-1.950	2.340	0.000	0.871	0.759	0.282	-0.318
3	475	4.710	-2.170	2.540	-0.002	0.896	0.802	0.559	0.140
4	473	4.060	-1.730	2.330	0.001	0.791	0.626	0.284	0.063
7	476	4.240	-1.840	2.390	-0.001	0.847	0.717	0.347	-0.277
8	475	3.910	-1.700	2.200	0.000	0.865	0.749	0.127	-0.445
9	473	3.980	-1.820	2.160	-0.002	0.803	0.645	0.263	-0.351
10	471	4.740	-2.040	2.710	-0.003	0.889	0.790	0.413	-0.046
12	478	5.090	-2.340	2.760	-0.001	0.905	0.819	0.342	-0.265
13	487	4.080	-1.830	2.250	0.000	0.840	0.705	0.252	-0.586
รวม	**4,292**	**5.093**	**-2.337**	**2.756**	**-0.001**	**0.856**	**0.733**	**0.325**	**-0.216**

ผลการแปลงค่าคะแนนสเกลการประเมินผลการรู้เรื่องวิทยาศาสตร์ พบว่า คะแนน การประเมินผลการรู้เรื่องวิทยาศาสตร์ก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันมีคะแนนเฉลี่ยสูงกว่าหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน และคะแนนการประเมินผลการรู้เรื่องวิทยาศาสตร์ก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันในภาพรวมทุกฉบับมีคะแนนเฉลี่ยเท่ากับ 484.710 คะแนน มากที่สุดเท่ากับ 767.568 และคะแนนน้อยที่สุดเท่ากับ 278.944 ส่วนคะแนนการประเมินผลการรู้เรื่องวิทยาศาสตร์หลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันในภาพรวมทุกฉบับมีคะแนนเฉลี่ยเท่ากับ 484.581 คะแนนมากที่สุดเท่ากับ 741.645 และน้อยที่สุดเท่ากับ 266.729 รายละเอียดตามตารางที่ 44

ตารางที่ 44 สรุปผลการแปลงค่าคะแนนสเกลการประเมินผลการรู้เรื่องวิทยาศาสตร์

ฉบับที่	N	คะแนนก่อนตัดข้อสอบที่เกิด DIF				คะแนนหลังตัดข้อสอบที่เกิด DIF			
		Mean	Min	Max	SD	Mean	Min	Max	SD
2	484	484.534	309.157	676.278	79.663	484.614	302.536	702.480	81.236
3	475	484.448	297.408	744.349	85.331	484.492	282.581	721.783	83.512
4	473	484.910	310.649	731.574	77.588	484.696	323.331	701.734	73.757
7	476	484.794	301.231	721.410	82.170	484.594	312.700	707.702	78.950
8	475	484.679	320.440	715.535	81.753	484.641	325.755	689.892	80.679
9	473	484.776	298.247	717.120	77.737	484.505	315.311	686.162	74.911
10	471	484.567	295.449	739.873	83.969	484.406	294.703	737.076	82.877
12	478	484.791	278.944	767.568	86.493	484.582	266.729	741.645	84.400
13	487	484.875	297.408	707.516	81.517	484.691	313.726	694.181	78.313
รวม	**4,292**	**484.710**	**278.944**	**767.568**	**81.779**	**484.581**	**266.729**	**741.645**	**79.854**

ผลการเปรียบเทียบคะแนนเฉลี่ยการประเมินผลการรู้เรื่องวิทยาศาสตร์ของตัวแปรที่นำมาศึกษาการทำหน้าที่ต่างกันของข้อสอบก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน จำนวน 5 ตัวแปร ได้แก่ เพศ การเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา เศรษฐานะของครอบครัว ทรัพยากรของครอบครัว และความมั่งคั่งของครอบครัว ซึ่งตัวแปรเพศก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า เพศหญิงมีเฉลี่ยคะแนนเฉลี่ยการประเมินผลการรู้เรื่องวิทยาศาสตร์สูงกว่าเพศชายอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนตัวแปรการเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษาก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า นักเรียนที่เรียนพิเศษวิทยาศาสตร์นอกสถานศึกษามีคะแนนเฉลี่ยการประเมินผลการรู้เรื่องวิทยาศาสตร์สูงกว่านักเรียนที่ไม่เรียนพิเศษวิทยาศาสตร์นอกสถานศึกษาอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ตัวแปรเศรษฐานะของครอบครัวก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า นักเรียนที่มีเศรษฐานะของครอบครัวต่ำมีคะแนนเฉลี่ยการประเมินผลการรู้เรื่องวิทยาศาสตร์สูงกว่านักเรียนที่มีเศรษฐานะของครอบครัวสูงอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ตัวแปรแหล่งทรัพยากรการศึกษาที่บ้านก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า นักเรียนที่มีแหล่งทรัพยากรการศึกษาที่บ้านสูงมีคะแนนเฉลี่ยการประเมินผลการรู้เรื่องวิทยาศาสตร์สูงกว่านักเรียนที่มีแหล่งทรัพยากรการศึกษาที่บ้านต่ำอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และตัวแปรความมั่งคั่งของครอบครัวก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า นักเรียนที่มีความมั่งคั่งของครอบครัวสูงมีคะแนนเฉลี่ยการประเมินผลการรู้เรื่องวิทยาศาสตร์สูงกว่านักเรียนที่มีความมั่งคั่งของครอบครัวต่ำอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 รายละเอียดตามตารางที่ 45

ตารางที่ 45 เปรียบเทียบคะแนนเฉลี่ยของตัวแปรก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน

ตัวแปร			N	Mean	SD	Levene's Test of Eqaual of Variance			t	df	p
						F	Sig.	Equal variances			
เพศ	ก่อนตัดข้อสอบ DIF	หญิง	2439	489.25	80.64	2.47	.12	assumed	4.18*	4290.00	.00
		ชาย	1853	478.73	82.89						
	หลังตัดข้อสอบ DIF	หญิง	2439	487.49	79.03	2.17	.14	assumed	2.74*	4290.00	.01
		ชาย	1853	480.75	80.79						
การเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา	ก่อนตัดข้อสอบ DIF	เรียน	1279	517.56	88.37	25.72	.00	not assumed	-7.73*	2359.00	.00
		ไม่เรียน	1085	491.18	77.64						
	หลังตัดข้อสอบ DIF	เรียน	1279	515.96	86.49	22.43	.00	not assumed	-7.55*	2357.00	.00
		ไม่เรียน	1085	490.59	76.91						
เศรษฐานะของครอบครัว	ก่อนตัดข้อสอบ DIF	สูง	965	542.77	84.41	36.20	.00	not assumed	-24.95*	1408.00	.00
		ต่ำ	3316	467.99	72.81						
	หลังตัดข้อสอบ DIF	สูง	965	540.44	82.62	40.31	.00	not assumed	-24.52*	1408.00	.00
		ต่ำ	3316	468.50	71.27						
แหล่งทรัพยากรการศึกษาที่บ้าน	ก่อนตัดข้อสอบ DIF	สูง	1736	517.22	85.35	83.00	.00	not assumed	-21.93*	3272.00	.00
		ต่ำ	2550	462.65	71.30						
	หลังตัดข้อสอบ DIF	สูง	1736	515.50	83.74	93.50	.00	not assumed	-21.28*	3265.00	.00
		ต่ำ	2550	463.62	69.73						
ความมั่งคั่งของครอบครัว	ก่อนตัดข้อสอบ DIF	สูง	624	538.93	88.21	23.02	.00	not assumed	-16.90*	792.79	.00
		ต่ำ	3663	475.49	76.96						
	หลังตัดข้อสอบ DIF	สูง	624	536.59	86.87	29.04	.00	not assumed	-16.48*	789.76	.00
		ต่ำ	3663	475.75	75.15						

หมายเหตุ * p<.05

ตอนที่ 3 ผลการตรวจสอบเงื่อนไขการวิเคราะห์พหุระดับและความเป็นเอกมิติของแบบสอบ

ในส่วนนี้เกี่ยวข้องกับการวิเคราะห์พหุระดับ เพื่อให้การวิเคราะห์มีความถูกต้องและเป็นไปตามเงื่อนไข ผู้วิจัยจึงได้แบ่งเนื้อหาในการนำเสนอออกเป็น 3 ตอนย่อย ได้แก่ ตอนย่อยที่หนึ่ง คือ การตรวจสอบเงื่อนไขการวิเคราะห์พหุระดับ ตอนย่อยที่สอง คือ โมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ และตอนย่อยที่สาม คือ การวิเคราะห์โมเดลมูลค่าเพิ่มพหุระดับ รายละเอียดดังนี้

3.1 การตรวจสอบเงื่อนไขการวิเคราะห์พหุระดับ

การตรวจสอบเงื่อนไขการวิเคราะห์พหุระดับในเบื้องต้นคือการตรวจสอบความสัมพันธ์ระหว่างตัวแปรโดยการตรวจสอบภาวะร่วมเส้นตรงพหุ (multicollinearity) และการตรวจสอบความสอดคล้องกับสมการการวิเคราะห์ถดถอยและโมเดลพหุระดับ นอกจากนี้ผู้วิจัยได้กำหนด

สัญลักษณ์และความหมายของตัวแปรระดับนักเรียนและสถานศึกษาเพื่อความสะดวกและง่ายต่อการทำความเข้าใจ ดังนี้

ระดับตัวแปร	ชื่อตัวแปร	ความหมาย
ระดับนักเรียน	SCIE1	คะแนนสเกลประเมินผลการรู้เรื่องวิทยาศาสตร์ก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน
	SCIE2	คะแนนสเกลประเมินผลการรู้เรื่องวิทยาศาสตร์หลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน
	FEMALE	ตัวแปรดัมมี่ความเป็นเพศหญิงของนักเรียน
	EXTRA	ตัวแปรดัมมี่การใช้เวลาเรียนในการเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา
	ESCS	เศรษฐานะของครอบครัว
	HEDRES	แหล่งทรัพยากรการศึกษาที่บ้าน
	WEALTH	ความมั่งคั่งของครอบครัว
	ICTRES	แหล่งทรัพยากรทางไอซีที
ระดับสถานศึกษา	PUBLIC	ตัวแปรดัมมี่สถานศึกษาของรัฐ
	TCSHORT	ระดับการขาดแคลนครู
	FUNGOV	ร้อยละงบประมาณของรัฐที่จัดสรรให้สถานศึกษา
	SCMATEDU	ทรัพยากรการเรียน
	STRATIO	สัดส่วนนักเรียนต่อครู
	RESPCURR	หลักสูตรการวัดและประเมินผล

3.1.1 การตรวจสอบภาวะร่วมเส้นตรงพหุ

การตรวจสอบภาวะร่วมเส้นตรงพหุ เป็นการตรวจสอบความสัมพันธ์ระหว่างตัวแปรทำนายคู่ที่มีความสัมพันธ์กันสูงหรือภาวะร่วมเส้นตรงพหุ (multicollinearity) ถ้าหากพบตัวแปรต้นคู่ที่มีความสัมพันธ์กันสูงมากขึ้นจะทำให้เพิ่มโอกาสของความคลาดเคลื่อนประเภทที่ 2 (Type II error) มากขึ้น (Field & Andy, 2003 อ้างถึงใน บุญเรือง ขจรศิลป์, 2555) ซึ่งจะต้องเลือกตัวแปรใดตัวแปรหนึ่งเข้าสู่สมการการวิเคราะห์แทน

การวิเคราะห์ค่าสหสัมพันธ์ของตัวแปรในระดับนักเรียนก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน ได้แก่ คะแนนสเกลประเมินผลการรู้เรื่องวิทยาศาสตร์ก่อนตัดข้อสอบที่ตรวจพบ

การทำหน้าที่ต่างกัน (SCIE1) ตัวแปรดัมมี่ความเป็นเพศหญิงของนักเรียน (FEMALE) ตัวแปรดัมมี่การใช้เวลาเรียนในการเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา (EXTRA) เศรษฐานะของครอบครัว (ESCS) แหล่งทรัพยากรการศึกษาที่บ้าน (HEDRES) ความมั่งคั่งของครอบครัว (WEALTH) และแหล่งทรัพยากรทางไอซีที (ICTRES) เมื่อพิจารณาความสัมพันธ์ของตัวแปรระดับนักเรียน พบว่า ตัวแปรตาม คือ คะแนนสเกลประเมินผลการรู้เรื่องวิทยาศาสตร์ก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน (SCIE1) มีความสัมพันธ์กับตัวแปรทำนายทั้ง 6 ตัวแปรอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 เรียงลำดับตามขนาดความสัมพันธ์ ได้แก่ เศรษฐานะของครอบครัว (ESCS) ($r = .401, p<.01$), แหล่งทรัพยากรการศึกษาที่บ้าน (HEDRES) ($r = .375, p<.01$), แหล่งทรัพยากรทางไอซีที (ICTRES) ($r = .364, p<.01$), ความมั่งคั่งของครอบครัว (WEALTH) ($r = .300, p<.01$), ตัวแปรดัมมี่การใช้เวลาเรียนในการเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา (EXTRA) ($r = .155, p<.01$), และตัวแปรดัมมี่ความเป็นเพศหญิงของนักเรียน (FEMALE) ($r = .064, p<.01$) ตามลำดับ ส่วนความสัมพันธ์ระหว่างตัวแปรทำนายที่มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และ .05 พบว่า แหล่งทรัพยากรทางไอซีทีกับความมั่งคั่งของครอบครัวมีความสัมพันธ์สูงที่สุด ($r_{ICTRES, WEALTH} = .795, p<.01$) รองลงมาคือ แหล่งทรัพยากรทางไอซีทีกับแหล่งทรัพยากรการศึกษาที่บ้าน ($r_{ICTRES, HEDRES} = .787, p<.01$) และความมั่งคั่งของครอบครัวและเศรษฐานะของครอบครัว ($r_{WEALTH, ESCS} = .783, p<.01$) ตามลำดับ รายละเอียดตามตารางที่ 46

ตารางที่ 46 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรระดับนักเรียนก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน

ตัวแปร	SCIE1	FEMALE	EXTRA	ESCS	HEDRES	WEALTH	ICTRES
SCIE1	1.000						
FEMALE	.064**	1.000					
EXTRA	.155**	-0.03	1.000				
ESCS	.401**	-.030*	.197**	1.000			
HEDRES	.375**	.056**	.182**	.685**	1.000		
WEALTH	.300**	-.042**	.160**	.783**	.674**	1.000	
ICTRES	.364**	-0.016	.189**	.736**	.787**	.795**	1.000

หมายเหตุ * หมายถึง p< .05, ** หมายถึง p< .01

ส่วนการวิเคราะห์ค่าสหสัมพันธ์ของตัวแปรระดับนักเรียนหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน ได้แก่ คะแนนสเกลประเมินผลการรู้เรื่องวิทยาศาสตร์หลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน (SCIE2) ตัวแปรดัมมี่ความเป็นเพศหญิงของนักเรียน (FEMALE) ตัวแปรดัมมี่

การใช้เวลาเรียนในการเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา (EXTRA) เศรษฐานะของครอบครัว (ESCS) แหล่งทรัพยากรการศึกษาที่บ้าน (HEDRES) ความมั่งคั่งของครอบครัว (WEALTH) และแหล่งทรัพยากรทางไอซีที (ICTRES) เมื่อพิจารณาความสัมพันธ์ของตัวแปรระดับนักเรียน พบว่าตัวแปรตามคือ คะแนนสเกลประเมินผลการรู้เรื่องวิทยาศาสตร์หลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน (SCIE2) มีความสัมพันธ์กับตัวแปรทำนายทั้ง 6 ตัวแปร อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 เรียงลำดับตามขนาดความสัมพันธ์ ได้แก่ เศรษฐานะของครอบครัว (ESCS) ($r = .394$, $p<.01$), แหล่งทรัพยากรการศึกษาที่บ้าน (HEDRES) ($r = .366$, $p<.01$), แหล่งทรัพยากรทางไอซีที (ICTRES) ($r = .356$, $p<.01$), ความมั่งคั่งของครอบครัว (WEALTH) ($r = .292$, $p<.01$), ตัวแปรดัมมี่การใช้เวลาเรียนในการเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา (EXTRA) ($r = .152$, $p<.01$), และตัวแปรดัมมี่ความเป็นเพศหญิงของนักเรียน (FEMALE) ($r = .042$, $p<.01$) ตามลำดับ ส่วนความสัมพันธ์ระหว่างตัวแปรทำนายที่มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และ .05 พบว่า แหล่งทรัพยากรทางไอซีทีกับความมั่งคั่งของครอบครัวมีความสัมพันธ์สูงที่สุด ($r_{ICTRES, WEALTH} = .795$, $p<.01$) รองลงมาคือ แหล่งทรัพยากรทางไอซีทีกับแหล่งทรัพยากรการศึกษาที่บ้าน ($r_{ICTRES, HEDRES} = .787$, $p<.01$) และความมั่งคั่งของครอบครัวและเศรษฐานะของครอบครัว ($r_{WEALTH, ESCS} = .783$, $p<.01$) ตามลำดับ รายละเอียดตามตารางที่ 47

ตารางที่ 47 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรระดับนักเรียนหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน

ตัวแปร	SCIE2	FEMALE	EXTRA	ESCS	HEDRES	WEALTH	ICTRES
SCIE2	1.000						
FEMALE	.042**	1.000					
EXTRA	.152**	-0.03	1.000				
ESCS	.394**	-.030*	.197**	1.000			
HEDRES	.366**	.056**	.182**	.685**	1.000		
WEALTH	.292**	-.042**	.160**	.783**	.674**	1.000	
ICTRES	.356**	-0.016	.189**	.736**	.787**	.795**	1.000

หมายเหตุ * หมายถึง $p<.05$, ** หมายถึง $p<.01$

การตรวจสอบภาวะร่วมเส้นตรงพหุ (multicollinearity) ก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันในตัวแปรระดับนักเรียน ผู้วิจัยได้วิเคราะห์ถดถอยพหุ (multiple regression) มีตัวแปรตามเป็นคะแนนสเกลประเมินผลการรู้เรื่องวิทยาศาสตร์ก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน และตัวแปรทำนาย ได้แก่ ตัวแปรดัมมี่ความเป็นเพศหญิง

ของนักเรียน (FEMALE) ตัวแปรดัมมี่การใช้เวลาเรียนในการเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา (EXTRA) เศรษฐานะของครอบครัว (ESCS) แหล่งทรัพยากรการศึกษาที่บ้าน (HEDRES) ความมั่งคั่งของครอบครัว (WEALTH) และแหล่งทรัพยากรทางไอซีที (ICTRES) โดยพิจารณาค่าความคลาดเคลื่อนยินยอม (Tolerance) และค่า VIF (variance inflation factor) ซึ่งค่าความคลาดเคลื่อนยินยอม (Tolerance) ควรสูงกว่า 0.1 และค่า VIF ควรมีค่าน้อยกว่า 10 (Hair et al., 1998) ส่วนค่าดัชนีเงื่อนไข (condition index) ไม่ควรเกิน 30 (สมถวิล วิจิตรวรรณา, สุภมาส อังศุโชติ และรัชนีกูล ภิญโญภานุวัฒน์, 2553) จากการตรวจสอบ พบว่า ก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันของตัวแปรทำนายทั้ง 6 ตัวแปร มีค่าความคลาดเคลื่อนยินยอม (Tolerance) สูงกว่า 0.1 และค่า VIF น้อยกว่า 10 ส่วนค่าดัชนีเงื่อนไข (condition index) ทั้ง 6 ตัวแปร ไม่เกิน 30 นั้นคือ ไม่เกิดปัญหาตัวแปรทำนายมีความสัมพันธ์กันสูงหรือไม่เกิดภาวะร่วมเส้นตรงพหุ ดังนั้นจึงไม่ต้องตัดตัวแปรทำนายระดับนักเรียนตัวใดทิ้ง รายละเอียดตามตารางที่ 48

ตารางที่ 48 ค่า Tolerance และ VIF จากผลการทดสอบความสัมพันธ์ระหว่างตัวแปรทำนายระดับนักเรียนก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน

ตัวแปร	ก่อนตัดข้อสอบ DIF			หลังตัดข้อสอบ DIF		
	condition index	collinearity Statistics		condition index	collinearity Statistics	
		Tolerance	VIF		Tolerance	VIF
FEMALE	1.495	0.986	1.014	1.495	0.986	1.014
EXTRA	3.039	0.955	1.047	3.039	0.955	1.047
ESCS	3.804	0.310	3.222	3.804	0.310	3.222
HEDRES	4.833	0.345	2.895	4.833	0.345	2.895
WEALTH	5.388	0.262	3.822	5.388	0.262	3.822
ICTRES	6.759	0.229	4.376	6.759	0.229	4.376

ผลการวิเคราะห์ค่าสหสัมพันธ์ของตัวแปรในระดับสถานศึกษาก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน ผู้วิจัยใช้คะแนนเฉลี่ยสเกลประเมินผลการรู้เรื่องวิทยาศาสตร์ก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน (SCIE1) เป็นตัวแปรตามในระดับสถานศึกษา ส่วนตัวแปรทำนายได้แก่ ตัวแปรดัมมี่สถานศึกษาของรัฐ (PUBLIC) ระดับการขาดแคลนครู (TCSHORT) ร้อยละงบประมาณของรัฐที่จัดสรรให้สถานศึกษา (FUNGOV) ทรัพยากรการเรียน (SCMATEDU) สัดส่วนนักเรียนต่อครู (STRATIO) และหลักสูตรการวัดและประเมินผล (RESPCURR) พบว่า ตัวแปรตามคือ คะแนนสเกลประเมินผลการรู้เรื่องวิทยาศาสตร์ (SCIE1) มีความสัมพันธ์กับตัวแปรทำนาย 2 ตัวแปร อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ได้แก่ ร้อยละงบประมาณของรัฐที่จัดสรรให้

สถานศึกษา (FUNGOV) (r = -.411, p<.01) มีทิศทางตรงกันข้าม ส่วนทรัพยากรการเรียน (SCMATEDU) (r = .326, p<.01) มีทิศทางเดียวกัน ส่วนความสัมพันธ์ระหว่างตัวแปรทำนายที่มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และ .05 พบว่า ทรัพยากรการเรียนกับระดับการขาดแคลนครูความสัมพันธ์สูงที่สุดในทิศทางตรงกันข้าม ($r_{SCMATEDU, TCSHORT}$ = -.431, p<.01) รองลงมาคือ ทรัพยากรการเรียนกับร้อยละงบประมาณของรัฐที่จัดสรรให้สถานศึกษาในทิศทางตรงกันข้ามเช่นเดียวกัน ($r_{SCMATEDU, FUNGOV}$ = -.403, p<.01) และร้อยละงบประมาณของรัฐที่จัดสรรให้สถานศึกษากับตัวแปรดัมมี่สถานศึกษาของรัฐ ($r_{FUNGOV, PUBLIC}$ = .255, p<.01) มีทิศทางเดียวกัน รายละเอียดตามตารางที่ 49

ตารางที่ 49 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรระดับสถานศึกษาก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน

ตัวแปร	SCIE1	PUBLIC	TCSHORT	FUNGOV	SCMATEDU	STRATIO	RESPCURR
SCIE1	1.000						
PUBLIC	0.111	1.000					
TCSHORT	-0.061	.149*	1.000				
FUNGOV	-.411**	.255**	0.102	1.000			
SCMATEDU	.326**	-.248**	-.431**	-.403**	1.000		
STRATIO	-0.065	-0.024	0.006	0.012	0.047	1.000	
RESPCURR	0.091	-0.04	-0.063	-.173**	0.024	-.194**	1.000

หมายเหตุ * หมายถึง p<.05, ** หมายถึง p<.01

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ส่วนผลการวิเคราะห์ค่าสหสัมพันธ์ของตัวแปรในระดับสถานศึกษาหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน ผู้วิจัยใช้คะแนนเฉลี่ยสเกลประเมินผลการรู้เรื่องวิทยาศาสตร์หลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน (SCIE2) เป็นตัวแปรตามในระดับสถานศึกษา ส่วนตัวแปรทำนายได้แก่ ตัวแปรดัมมี่สถานศึกษาของรัฐ (PUBLIC) ระดับการขาดแคลนครู (TCSHORT) ร้อยละงบประมาณของรัฐที่จัดสรรให้สถานศึกษา (FUNGOV) ทรัพยากรการเรียน (SCMATEDU) สัดส่วนนักเรียนต่อครู (STRATIO) และหลักสูตรการวัดและประเมินผล (RESPCURR) พบว่า ตัวแปรตามคือ คะแนนสเกลประเมินผลการรู้เรื่องวิทยาศาสตร์ (SCIE2) มีความสัมพันธ์กับตัวแปรทำนาย 2 ตัวแปร อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ได้แก่ ร้อยละงบประมาณของรัฐที่จัดสรรให้สถานศึกษา (FUNGOV) (r = -.402, p<.01) มีทิศทางตรงกันข้าม ส่วนทรัพยากรการเรียน (SCMATEDU) (r = .308, p<.01) มีทิศทางเดียวกัน ส่วนความสัมพันธ์ระหว่างตัวแปรทำนายที่มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และ .05 พบว่า ทรัพยากรการเรียนกับ

ระดับการขาดแคลนครูความสัมพันธ์สูงที่สุดในทิศทางตรงกันข้าม ($r_{SCMATEDU, TCSHORT} = -.431$, $p<.01$) รองลงมาคือ ทรัพยากรการเรียนกับร้อยละงบประมาณของรัฐที่จัดสรรให้สถานศึกษาในทิศทางตรงกันข้ามเช่นเดียวกัน ($r_{SCMATEDU, FUNGOV} = -.403$, $p<.01$) และร้อยละงบประมาณของรัฐที่จัดสรรให้สถานศึกษากับตัวแปรดัมมี่สถานศึกษาของรัฐ ($r_{FUNGOV, PUBLIC} = .255$, $p<.01$) มีทิศทางเดียวกัน รายละเอียดตามตารางที่ 50

ตารางที่ 50 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรระดับสถานศึกษาหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน

ตัวแปร	SCIE2	PUBLIC	TCSHORT	FUNGOV	SCMATEDU	STRATIO	RESPCURR
SCIE2	1.000						
PUBLIC	0.108	1.000					
TCSHORT	-0.046	.149*	1.000				
FUNGOV	-.402**	.255**	0.102	1.000			
SCMATEDU	.308**	-.248**	-.431**	-.403**	1.000		
STRATIO	-0.087	-0.024	0.006	0.012	0.047	1.000	
RESPCURR	0.09	-0.04	-0.063	-.173**	0.024	-.194**	1.000

หมายเหตุ * หมายถึง $p<.05$, ** หมายถึง $p<.01$

การตรวจสอบภาวะร่วมเส้นตรงพหุ (multicollinearity) ก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันระดับสถานศึกษา ผู้วิจัยได้วิเคราะห์ถดถอยพหุ (multiple regression) มีตัวแปรตามเป็นคะแนนเฉลี่ยสเกลประเมินผลการรู้เรื่องวิทยาศาสตร์ก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน และตัวแปรทำนาย ได้แก่ ตัวแปรดัมมี่สถานศึกษาของรัฐ (PUBLIC) ระดับการขาดแคลนครู (TCSHORT) ร้อยละงบประมาณของรัฐที่จัดสรรให้สถานศึกษา (FUNGOV) ทรัพยากรการเรียน (SCMATEDU) สัดส่วนนักเรียนต่อครู (STRATIO) และหลักสูตรการวัดและประเมินผล (RESPCURR) โดยพิจารณาค่าความคลาดเคลื่อนยินยอม (Tolerance) และค่า VIF (variance inflation factor) ซึ่งค่าความคลาดเคลื่อนยินยอม (Tolerance) ควรสูงกว่า 0.1 และค่า VIF ควรมีค่าน้อยกว่า 10 (Hair et al., 1998) ส่วนค่าดัชนีเงื่อนไข (condition index) ไม่ควรเกิน 30 (สมถวิล วิจิตรวรรณา, สุภมาส อังศุโชติ และรัชนีกูล ภิญโญภานุวัฒน์, 2553) จากการตรวจสอบ พบว่า ก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันของตัวแปรทำนายทั้ง 6 ตัวแปร มีค่าความคลาดเคลื่อนยินยอม (Tolerance) สูงกว่า 0.1 และค่า VIF น้อยกว่า 10 ส่วนค่าดัชนีเงื่อนไข (condition index) ทั้ง 6 ตัวแปร ไม่เกิน 30 นั้นคือ ไม่เกิดปัญหาตัวแปรทำนาย

มีความสัมพันธ์กันสูงหรือไม่เกิดภาวะร่วมเส้นตรงพหุ ดังนั้นจึงไม่ต้องตัดตัวแปรทำนายระดับสถานศึกษาตัวใดทิ้ง รายละเอียดตามตารางที่ 51

ตารางที่ 51 ค่า Tolerance และ VIF จากผลการทดสอบความสัมพันธ์ระหว่างตัวแปรทำนายระดับสถานศึกษาก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน

ตัวแปร	ก่อนตัดข้อสอบ DIF			หลังตัดข้อสอบ DIF		
	condition index	Collinearity Statistics		condition index	Collinearity Statistics	
		Tolerance	VIF		Tolerance	VIF
PUBLIC	2.273	0.906	1.104	2.273	0.906	1.104
TCSHORT	3.201	0.800	1.250	3.201	0.800	1.250
FUNGOV	3.464	0.778	1.286	3.464	0.778	1.286
SCMATEDU	6.944	0.668	1.496	6.944	0.668	1.496
STRATIO	8.663	0.959	1.043	8.663	0.959	1.043
RESPCURR	14.871	0.927	1.078	14.871	0.927	1.078

ตอนที่ 4 ผลการเปรียบเทียบประสิทธิภาพโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์

ในส่วนผลการเปรียบเทียบประสิทธิภาพโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ ผู้วิจัยแบ่งการนำเสนอออกเป็น 5 ตอนย่อย ได้แก่ ตอนย่อยที่หนึ่ง คือ ลักษณะโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ ตอนย่อยที่สอง คือ การวิเคราะห์โมเดลมูลค่าเพิ่มพหุระดับ ตอนย่อยที่สาม คือ การสร้างเกณฑ์การตัดสินการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม ตอนย่อยที่สี่ คือ ความสัมพันธ์ระหว่างคะแนนเฉลี่ยการประเมินผลการรู้เรื่องวิทยาศาสตร์ คะแนนมูลค่าเพิ่ม การจัดอันดับคุณภาพคะแนนมูลค่าเพิ่ม และการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม และตอนย่อยสุดท้าย คือ การเปรียบเทียบการจัดกลุ่มคุณภาพการจัดการศึกษาวิทยาศาสตร์ระหว่าง 2 โมเดล

4.1 ลักษณะโมเดลการประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์

โมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ มีทั้งหมดจำนวน 4 โมเดล คือ เป็นโมเดลไม่มีเงื่อนไข (unconditional model หรือ null model) จำนวน 2 โมเดล คือโมเดลไม่มีเงื่อนไขที่ไม่ได้พิจารณาข้อสอบที่ทำหน้าที่ต่างกันข้อสอบและการทำหน้าที่ต่างกันของแบบสอบ และโมเดลไม่มีเงื่อนไขที่มีการพิจารณาข้อสอบที่ทำหน้าที่ต่างกันข้อสอบและการทำหน้าที่ต่างกันของแบบสอบ โดยที่ทั้ง 2 โมเดลไม่มีตัวแปรในระดับนักเรียนและสถานศึกษา และโมเดลสมมติฐานตามวัตถุประสงค์การวิจัยอีก 2 โมเดล คือ โมเดล Undetected DIF-DSF-DTF เป็นโมเดล

การประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ที่ไม่ได้พิจารณาการทำหน้าที่ต่างกันข้อสอบและแบบสอบ แต่มีการควบคุมอิทธิพลของตัวแปรในระดับนักเรียนและสถานศึกษา และโมเดล Detected DIF-DSF-DTF เป็นโมเดลการประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ที่พิจารณาการทำหน้าที่ต่างกันข้อสอบและแบบสอบ และมีการควบคุมตัวแปรในระดับนักเรียนและสถานศึกษา ในการวิเคราะห์ส่วนนี้ผู้วิจัยได้แสดงสมการวิเคราะห์ในแต่ละโมเดล ดังนี้

4.1.1 โมเดลไม่มีเงื่อนไข (unconditional model หรือ null model)

โมเดลไม่มีเงื่อนไข Undetected DIF-DSF-DTF เป็นโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ที่ไม่ได้พิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ และไม่มี การควบคุมตัวแปรในระดับนักเรียนและสถานศึกษา

ระดับที่ 1: ระดับนักเรียน (ระหว่างนักเรียน i ภายใต้สถานศึกษา j) (Student-level)

$SCIEACH1_{ij}=B_{0j} + R_{ij}$

ระดับที่ 2: ระดับหน่วย (ระหว่างสถานศึกษา) (School-level)

$B_{0j} = G_{00} + U_{0j}$

โมเดลไม่มีเงื่อนไข Detected DIF-DSF-DTF เป็นโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ที่พิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ และไม่มีการควบคุมตัวแปรในระดับนักเรียนและสถานศึกษา

ระดับที่ 1: ระดับนักเรียน (ระหว่างนักเรียน i ภายใต้สถานศึกษา j) (Student-level)

$SCIEACH2_{ij}=B_{0j} + R_{ij}$

ระดับที่ 2: ระดับหน่วย (ระหว่างสถานศึกษา) (School-level)

$B_{0j} = G_{00} + U_{0j}$

โดยที่

$SCIEACH1_{ij}$ = คะแนนการประเมินผลการรู้เรื่องวิทยาศาสตร์ที่ก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันของนักเรียนคนที่ i ภายในสถานศึกษา j

$SCIEACH2_{ij}$ = คะแนนการประเมินผลการรู้เรื่องวิทยาศาสตร์ที่ภายหลังข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันของนักเรียนคนที่ i ภายในสถานศึกษา j

B_{0j} = ค่าจุดตัดแกน (intercept) เป็นค่าคะแนนเฉลี่ยของผลการประเมินผลการรู้เรื่องวิทยาศาสตร์ในสถานศึกษา j

R_{ij} = ค่าเศษเหลือหรือค่าความคลาดเคลื่อน (residual error) ระดับนักเรียน

G_{00} = ค่าจุดตัด (intercept) เป็นค่าคะแนนเฉลี่ยของคะแนนผลการประเมินผลการรู้เรื่องวิทยาศาสตร์ในสถานศึกษาทุกสถานศึกษา

U_{0j} = ค่าเศษเหลือหรือค่าความคลาดเคลื่อน (residual error) ระดับหน่วย

4.1.2 โมเดลสมมติฐานตามวัตถุประสงค์การวิจัย

โมเดล Undetected DIF-DSF-DTF เป็นโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ที่ไม่ได้พิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ แต่ควบคุมตัวแปรของนักเรียนและสถานศึกษา โดยนำเข้ามาเป็นตัวแปรในการวิเคราะห์พหุระดับเพื่อหาเศษเหลือของสถานศึกษา ซึ่งมีโมเดลในการวิเคราะห์ดังต่อไปนี้

ระดับที่ 1: ระดับนักเรียน (ระหว่างนักเรียน i ภายใต้สถานศึกษา j) (Student-level)

$$SCIEACH1_{ij}=B_{0j} + B_{1j}*(FEMALE)_{ij} + B_{2j}*(EXTRA)_{ij} + B_{3j}*(ESCS)_{ij} + B_{4j}*(HEDRES)_{ij} + B_{5j}*(WEALTH)_{ij} + B_{6j}*(ICTRES)_{ij} + R_{ij}$$

ระดับที่ 2: ระดับหน่วย (ระหว่างสถานศึกษา) (School-level)

$$B_{0j} = G_{00} + G_{01}*(PUBLIC) + G_{02}*(TCSHORT) + G_{03}*(FUNGOV) + G_{04}*(SCMATEDU) + G_{05}*(STRATIO) + G_{06}*(RESPCURR) + U_{0j}$$

$$B_{1j} = G_{10}$$

$$B_{2j} = G_{20}$$

$$B_{3j} = G_{30}$$

$$B_{4j} = G_{40}$$

$$B_{5j} = G_{50}$$

$$B_{6j} = G_{60}$$

สมการผสมในโมเดลสมมติฐาน (Equation Mixed Hypothetical Model)

$$SCIEACH1_{ij} = G_{00} + G_{01}*(PUBLIC) + G_{02}*(TCSHORT - \overline{TCSHORT}) + G_{03}*(FUNGOV - \overline{FUNGOV}) + G_{04}*(SCMATEDU - \overline{SCMATEDU}) + G_{05}*(STRATIO - \overline{STRATIO}) + G_{06}*(RESPCURR - \overline{RESPCURR}) + G_{10}*(FEMALE) + G_{20}*(EXTRA) + G_{30}*(ESCS - \overline{ESCS}) + G_{40}*(HEDRES - \overline{HEDRES}) + G_{50}*(WEALTH - \overline{WEALTH}) + G_{60}*(ICTRES - \overline{ICTRES}) + R_{ij} + U_{0j}$$

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

โมเดล Detected DIF-DSF-DTF เป็นโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ที่พิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ และควบคุมตัวแปรของนักเรียนและสถานศึกษา โดยนำเข้ามาเป็นตัวแปรในการวิเคราะห์พหุระดับเพื่อหาเศษเหลือของสถานศึกษา ซึ่งมีโมเดลในการวิเคราะห์ดังต่อไปนี้

ระดับที่ 1: ระดับนักเรียน (ระหว่างนักเรียน i ภายใต้สถานศึกษา j)

$$SCIEACH2_{ij}=B_{0j} + B_{1j}*(FEMALE)_{ij} + B_{2j}*(EXTRA)_{ij} + B_{3j}*(ESCS)_{ij} + B_{4j}*(HEDRES)_{ij} + B_{5j}*(WEALTH)_{ij} + B_{6j}*(ICTRES)_{ij} + R_{ij}$$

ระดับที่ 2: ระดับหน่วย (ระหว่างสถานศึกษา)

$$B_{0j} = G_{00} + G_{01}*(PUBLIC) + G_{02}*(TCSHORT) + G_{03}*(FUNGOV) + G_{04}*(SCMATEDU) + G_{05}*(STRATIO) + G_{06}*(RESPCURR) + U_{0j}$$

$$B_{1j} = G_{10}$$

$$B_{2j} = G_{20}$$

$$B_{3j} = G_{30}$$

$$B_{4j} = G_{40}$$

$$B_{5j} = G_{50}$$

$$B_{6j} = G_{60}$$

สมการผสมในโมเดลสมมติฐาน (Equation Mixed Hypothetical Model)

$$SCIEACH2_{ij} = G_{00} + G_{01}*(PUBLIC) + G_{02}*(TCSHORT - \overline{TCSHORT}) + G_{03}*(FUNGOV - \overline{FUNGOV}) + G_{04}*(SCMATEDU - \overline{SCMATEDU}) + G_{05}*(STRATIO - \overline{STRATIO}) + G_{06}*(RESPCURR - \overline{RESPCURR}) + G_{10}*(FEMALE) + G_{20}*(EXTRA) + G_{30}*(ESCS - \overline{ESCS}) + G_{40}*(HEDRES - \overline{HEDRES}) + G_{50}*(WEALTH - \overline{WEALTH}) + G_{60}*(ICTRES - \overline{ICTRES}) + R_{ij} + U_{0j}$$

โดยที่

$SCIEACH1_{ij}$ = คะแนนการประเมินผลการรู้เรื่องวิทยาศาสตร์ที่ก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันของนักเรียนคนที่ i ภายในสถานศึกษา j

$SCIEACH2_{ij}$ = คะแนนการประเมินผลการรู้เรื่องวิทยาศาสตร์ที่ภายหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันของนักเรียนคนที่ i ภายในสถานศึกษา j

B_{0j} = ค่าจุดตัดแกน (intercept) เป็นค่าคะแนนเฉลี่ยของผลการประเมินผลการรู้เรื่องวิทยาศาสตร์ในสถานศึกษา j

R_{ij} = ค่าเศษเหลือหรือค่าความคลาดเคลื่อน (residual error)

G_{00} = ค่าจุดตัด (intercept) เป็นค่าคะแนนเฉลี่ยของคะแนนผลการประเมินผลการรู้เรื่องวิทยาศาสตร์ในสถานศึกษาทุกสถานศึกษา

G_{01} = ค่าความชัน (slope) ที่แสดงถึงค่าอิทธิพลของการเปลี่ยนแปลงคะแนนผลการประเมินการรู้เรื่องวิทยาวิทยาศาสตร์เมื่อมีการควบคุมตัวแปร Y_j

U_{0j} = ค่าเศษเหลือ เป็นค่าที่แสดงถึงคะแนนมูลค่าเพิ่มของสถานศึกษา

.

4.2 การวิเคราะห์โมเดลมูลค่าเพิ่มพหุระดับ

หลังจากการวิเคราะห์ข้อมูลด้วยโมเดลมูลค่าเพิ่มพหุระดับเป็นที่เรียบร้อยแล้ว ไฟล์ข้อมูลที่เรียกว่า residual file ในระดับที่ 2 ประกอบไปด้วยข้อมูลหลายส่วน แต่ส่วนที่สำคัญคือคะแนนมูลค่าเพิ่มของโรงเรียนแต่ละโรงเรียนที่ประมาณค่าแบบกำลังสองน้อยที่สุด (ordinary least square: OLS) และประมาณค่าแบบเบส์ (baye's estimation) ซึ่งผู้วิจัยเลือกใช้การประมาณค่าแบบเบส์ (Baye's Estimation) เพื่อนำคะแนนมูลค่าเพิ่ม (value-added scores or residual scores) มาเปรียบเทียบประสิทธิภาพระหว่างโมเดล Undetected DIF-DSF-DTF และโมเดล Detected DIF-DSF-DTF

ผลการวิเคราะห์โมเดลมูลค่าพหุระดับเพื่อนำคะแนนมูลค่าเพิ่มของสถานศึกษาเป็นตัววัดประสิทธิภาพของสถานศึกษาแต่ละโรงเรียน ซึ่งมีทั้งหมดจำนวน 4 โมเดล คือ เป็นโมเดลไม่มีเงื่อนไข (unconditional model หรือ null model) จำนวน 2 โมเดล คือโมเดลที่ไม่ได้พิจารณาการทำหน้าที่ต่างกันข้อสอบและแบบสอบ และโมเดลที่พิจารณาข้อสอบที่ทำหน้าที่ต่างกันข้อสอบและแบบสอบ โดยมีวัตถุประสงค์เพื่อตรวจสอบคะแนนสเกลการประเมินผลการรู้เรื่องวิทยาศาสตร์ที่ก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันแตกต่างจากศูนย์ และแตกต่างระหว่างโรงเรียนหรือไม่ ถ้าพบความผันแปรมีค่าต่างจากศูนย์แสดงว่าค่าพารามิเตอร์ B_{0j} หรือค่า intercept มีความผันแปรระหว่างโรงเรียนซึ่งในกรณีนี้โมเดลมูลค่าเพิ่มจึงมีความเหมาะสมที่จะใส่ตัวแปรทำนายในโมเดลเพื่ออธิบายความผันแปรดังกล่าว และโมเดลตามสมมติฐาน (hypothetical model) ที่ใส่ตัวแปรทำนายในโมเดล จำนวน 2 โมเดล ตามวัตถุประสงค์การวิจัย ได้แก่ โมเดลที่ Undetected DIF-DSF-DTF เป็นโมเดลการประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ที่ไม่ได้พิจารณาการทำหน้าที่ต่างกันข้อสอบและแบบสอบ แต่มีการควบคุมตัวแปรในระดับนักเรียนและสถานศึกษา และโมเดล Detected DIF-DSF-DTF เป็นโมเดลการประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ที่พิจารณาการทำหน้าที่ต่างกันข้อสอบและแบบสอบ และมีการควบคุมตัวแปรในระดับนักเรียนและสถานศึกษา สรุปผลการวิเคราะห์ในแต่ละโมเดลดังต่อไปนี้ (รายละเอียดผลการวิเคราะห์อยู่ในภาคผนวก ค)

ผลการวิเคราะห์โมเดลไม่มีเงื่อนไข (unconditional model หรือ null model) ใน null model 1 และ null model 2 จากผลการวิเคราะห์พหุระดับที่ไม่มีการใส่ตัวแปรระดับนักเรียนและสถานศึกษาใน null model 1 พบว่า ผลการทดสอบอิทธิพลคงที่ (fixed effect) มีค่าเฉลี่ยรวมคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์ทุกสถานศึกษาก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน (G_{00}) มีความผันแปรจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 (G_{00} = 492.674, t = 139.290, p = .000) และ null model 2 พบว่า ผลการทดสอบอิทธิพลคงที่ (Fixed effect) มีค่าเฉลี่ยรวมคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์ทุกสถานศึกษาหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน (G_{00}) มีความผันแปรจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 (G_{00} = 492.246, t = 144.122, p = .000) ทั้ง 2 โมเดลมีความผันแปรระหว่างโรงเรียนซึ่งมีความเหมาะสมที่จะใส่ตัวแปรทำนายในโมเดลเพื่ออธิบายความผันแปรซึ่งในโมเดลมูลค่าเพิ่ม

เมทริกของความแปรปรวนร่วมหรือสหสัมพันธ์ภายในชั้น (intra-class correlation) ของ null model 1 เท่ากับ 0.318 หมายความว่า 31.8% เป็นความแปรปรวนของคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์ระหว่างโรงเรียน อีก 68.2% เป็นความแปรปรวนของคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์ระหว่างนักเรียน การประมาณค่าความเที่ยงของสัมประสิทธิ์ตัวแปรสุ่มในระดับที่ 1 มีค่าเท่ากับ 0.778 แสดงว่าค่าเฉลี่ยคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์มีค่าความเที่ยงตามเกณฑ์ประเมิน ของ George และ Mallery (2003) ระดับพอใช้ โดยมีฟังก์ชั่นล็อคไลค์ลิฮูด เท่ากับ -13,489.95 ส่วนเมทริกของความแปรปรวนร่วมหรือสหสัมพันธ์ภายในชั้น (intra-class correlation) ของ null model 2 เท่ากับ 0.302 หมายความว่า 30.2% เป็นความแปรปรวนของคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์ระหว่างโรงเรียน อีก 69.8% เป็นความแปรปรวนของคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์ระหว่างนักเรียน การประมาณค่าความเที่ยงของสัมประสิทธิ์ตัวแปรสุ่มในระดับที่ 1 มีค่าเท่ากับ 0.766 แสดงว่าค่าเฉลี่ยคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์มีค่าความเที่ยงตามเกณฑ์ประเมิน ของ George และ Mallery (2003) ระดับพอใช้ โดยมีฟังก์ชั่นล็อคไลค์ลิฮูด เท่ากับ -13,471.09

โมเดลสมมติฐานตามวัตถุประสงค์การวิจัยในโมเดล Undetected DIF-DSF-DTF (model 1) จากผลการวิเคราะห์พหุระดับที่มีการควบคุมตัวแปรระดับนักเรียนและสถานศึกษา ผลการทดสอบอิทธิพลคงที่ (fixed effect) พบว่า ค่าเฉลี่ยรวมคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์ทุกสถานศึกษา (G_{00}) มีความผันแปรจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 (G_{00} = 465.762, t = 58.115, p = .000) ซึ่งหมายความว่าค่าเฉลี่ยคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์ของนักเรียนทุกคน (grand mean) ไม่เป็นศูนย์หรือแตกต่างจากศูนย์ โดยมีค่าสัมประสิทธิ์ถดถอยของตัวแปรระดับสถานศึกษาที่พบว่ามีนัยสำคัญทางสถิติที่ระดับ .01 ที่มีค่าเป็น

บวกสูงที่สุด คือ สถานศึกษาของรัฐบาลมีค่าเป็น 36.595 หมายความว่า ถ้าเป็นโรงเรียนรัฐบาลเพิ่มขึ้น 1 หน่วยมาตรฐาน (1SD) จะมีผลต่อการเพิ่มขึ้นของคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์ 36.595 หน่วย เมื่อควบคุมตัวแปรทำนายตัวอื่นๆ ที่อยู่ในสมการทำนายให้คงที่ ด้วยความเชื่อมั่น 99% รองลงมาคือ ทรัพยากรการเรียนมีค่าเป็น 12.413 หมายความว่าถ้าโรงเรียนมีทรัพยากรการเรียนเพิ่มขึ้น 1 หน่วยมาตรฐาน (1SD) จะมีผลต่อการเพิ่มขึ้นของคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์ 12.413 หน่วย เมื่อควบคุมตัวแปรทำนายตัวอื่นๆ ที่อยู่ในสมการทำนายให้คงที่ ด้วยความเชื่อมั่น 99% ส่วนค่าสัมประสิทธิ์ถดถอยของตัวแปรระดับสถานศึกษาที่พบว่ามีนัยสำคัญทางสถิติที่ระดับ .01 ที่มีค่าเป็นลบสูงที่สุด คือ ร้อยละงบประมาณของรัฐที่จัดสรรให้สถานศึกษามีค่าเป็น -0.414 หมายความว่า ถ้าโรงเรียนมีร้อยละงบประมาณของที่จัดสรรให้สถานศึกษาเพิ่มขึ้น 1 หน่วยมาตรฐาน (1SD) จะมีผลต่อการลดดงของคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์ 0.414 หน่วย เมื่อควบคุมตัวแปรทำนายตัวอื่นๆ ที่อยู่ในสมการทำนายให้คงที่ ด้วยความเชื่อมั่น 99%

ส่วนสัมประสิทธิ์ถดถอยของตัวแปรระดับนักเรียนที่พบว่ามีนัยสำคัญทางสถิติที่ระดับ .01 ที่มีค่าเป็นบวกสูงที่สุด คือ เศรษฐานะของครอบครัวมีค่าเป็น 13.183 หมายความว่า นักเรียนที่มีเศรษฐานะของครอบครัวเพิ่มขึ้น 1 หน่วยมาตรฐาน (1SD) จะมีผลต่อการเพิ่มขึ้นของคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์ 13.183 หน่วย เมื่อควบคุมตัวแปรทำนายตัวอื่นๆ ที่อยู่ในสมการทำนายให้คงที่ ด้วยความเชื่อมั่น 99% รองลงมาคือ แหล่งทรัพยากรการศึกษาที่บ้านมีค่าเป็น 10.709 หมายความว่า นักเรียนที่มีแหล่งทรัพยากรการศึกษาที่บ้านเพิ่มขึ้น 1 หน่วยมาตรฐาน (1SD) จะมีผลต่อการเพิ่มขึ้นของคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์ 10.709 หน่วย เมื่อควบคุมตัวแปรทำนายตัวอื่นๆ ที่อยู่ในสมการทำนายให้คงที่ ด้วยความเชื่อมั่น 99% ส่วนสัมประสิทธิ์ถดถอยของตัวแปรระดับนักเรียนที่พบว่ามีนัยสำคัญทางสถิติที่ระดับ .01 ที่มีค่าเป็นลบสูงที่สุดคือ ความมั่งคั่งของครอบครัวมีค่าเป็น -16.088 หมายความว่า นักเรียนที่มีความมั่งคั่งของครอบครัวเพิ่มขึ้น 1 หน่วยมาตรฐาน (1SD) จะมีผลต่อการลดลงของคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์ 16.088 หน่วย เมื่อควบคุมตัวแปรทำนายตัวอื่นๆ ที่อยู่ในสมการทำนายให้คงที่ ด้วยความเชื่อมั่น 99%

ผลการทดสอบอิทธิพลเชิงสุ่ม (random effect) พบว่า ค่าเศษเหลือระดับสถานศึกษาของคะแนนการประเมินผลการรู้เรื่องวิทยาศาสตร์ที่มีการควบคุมตัวแปรระดับนักเรียนและสถานศึกษา (U_{0k}) หรือมูลค่าเพิ่มของสถานศึกษาที่ความผันแปรระหว่างสถานศึกษาอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 (χ^2 = 684.761, p = .000) เมื่อพิจารณาเมทริกของความแปรปรวนร่วมหรือสหสัมพันธ์ภายในชั้น (intra-class correlation) ของโมเดลเท่ากับ 0.161 หมายความว่า 16.1% เป็น

ความแปรปรวนของคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์ระหว่างโรงเรียน อีก 83.9% เป็นความแปรปรวนของคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์ระหว่างนักเรียน การประมาณค่าความเที่ยงของสัมประสิทธิ์ตัวแปรสุ่มในระดับที่ 1 มีค่าเท่ากับ 0.614 แสดงว่าค่าเฉลี่ยคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์มีค่าความเที่ยงตามเกณฑ์ประเมิน ของ George และ Mallery (2003) ระดับค่อนข้างพอใช้ โดยมีฟังก์ชั่นล็อคไลค์ลิฮูด เท่ากับ -13,370.56

โมเดลสมมติฐานตามวัตถุประสงค์การวิจัยในโมเดล Detected DIF-DSF-DTF (model 2) จากผลการวิเคราะห์พหุระดับที่มีการควบคุมตัวแปรระดับนักเรียนและสถานศึกษา ผลการทดสอบอิทธิพลคงที่ (fixed effect) พบว่า ค่าเฉลี่ยรวมคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์ทุกสถานศึกษา (G_{00}) มีความผันแปรจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 (G_{00} = 469.839, t = 60.135, p = .000) ซึ่งหมายความว่าค่าเฉลี่ยคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์ของนักเรียนทุกคน (grand mean) ไม่เป็นศูนย์หรือแตกต่างจากศูนย์ โดยมีค่าสัมประสิทธิ์ถดถอยของตัวแปรระดับสถานศึกษาที่พบว่ามีนัยสำคัญทางสถิติที่ระดับ .01 ที่มีค่าเป็นบวกสูงที่สุด คือ สถานศึกษาของรัฐบาลมีค่าเป็น 33.688 หมายความว่า ถ้าเป็นโรงเรียนรัฐบาลเพิ่มขึ้น 1 หน่วยมาตรฐาน (1SD) จะมีผลต่อการเพิ่มขึ้นของคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์ 33.688 หน่วย เมื่อควบคุมตัวแปรทำนายตัวอื่นๆ ที่อยู่ในสมการทำนายให้คงที่ ด้วยความเชื่อมั่น 99% รองลงมาคือ ทรัพยากรการเรียนมีค่าเป็น 11.247 หมายความว่าถ้าโรงเรียนมีทรัพยากรการเรียนเพิ่มขึ้น 1 หน่วยมาตรฐาน (1SD) จะมีผลต่อการเพิ่มขึ้นของคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์ 11.247 หน่วย เมื่อควบคุมตัวแปรทำนายตัวอื่นๆ ที่อยู่ในสมการทำนายให้คงที่ ด้วยความเชื่อมั่น 99% ส่วนค่าสัมประสิทธิ์ถดถอยของตัวแปรระดับสถานศึกษาที่พบว่ามีนัยสำคัญทางสถิติที่ระดับ .01 ที่มีค่าเป็นลบสูงที่สุด คือ ร้อยละงบประมาณของรัฐที่จัดสรรให้สถานศึกษามีค่าเป็น -0.379 หมายความว่า ถ้าโรงเรียนมีร้อยละงบประมาณของที่จัดสรรให้สถานศึกษาเพิ่มขึ้น 1 หน่วยมาตรฐาน (1SD) จะมีผลต่อการลดดงของคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์ 0.379 หน่วย เมื่อควบคุมตัวแปรทำนายตัวอื่นๆ ที่อยู่ในสมการทำนายให้คงที่ ด้วยความเชื่อมั่น 99%

ส่วนสัมประสิทธิ์ถดถอยของตัวแปรระดับนักเรียนที่พบว่ามีนัยสำคัญทางสถิติที่ระดับ .01 ที่มีค่าเป็นบวกสูงที่สุด คือ เศรษฐานะของครอบครัวมีค่าเป็น 13.025 หมายความว่า นักเรียนที่มีเศรษฐานะของครอบครัวเพิ่มขึ้น 1 หน่วยมาตรฐาน (1SD) จะมีผลต่อการเพิ่มขึ้นของคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์ 13.025 หน่วย เมื่อควบคุมตัวแปรทำนายตัวอื่นๆ ที่อยู่ในสมการทำนายให้คงที่ ด้วยความเชื่อมั่น 99% รองลงมาคือ แหล่งทรัพยากรการศึกษาที่บ้านมีค่าเป็น 9.745 หมายความว่า นักเรียนที่มีแหล่งทรัพยากรการศึกษาที่บ้านเพิ่มขึ้น 1 หน่วยมาตรฐาน

(1SD) จะมีผลต่อการเพิ่มขึ้นของคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์ 9.745 หน่วย เมื่อควบคุมตัวแปรทำนายตัวอื่นๆ ที่อยู่ในสมการทำนายให้คงที่ ด้วยความเชื่อมั่น 99% ส่วนสัมประสิทธิ์ถดถอยของตัวแปรระดับนักเรียนที่พบว่ามีนัยสำคัญทางสถิติที่ระดับ .01 ที่มีค่าเป็นลบสูงที่สุดคือ ความมั่งคั่งของครอบครัวมีค่าเป็น -15.789 หมายความว่า นักเรียนที่มีความมั่งคั่งของครอบครัวเพิ่มขึ้น 1 หน่วยมาตรฐาน (1SD) จะมีผลต่อการลดลงของคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์ 15.789 หน่วย เมื่อควบคุมตัวแปรทำนายตัวอื่นๆ ที่อยู่ในสมการทำนายให้คงที่ ด้วยความเชื่อมั่น 99%

ผลการทดสอบอิทธิพลเชิงสุ่ม (random effect) พบว่า ค่าเศษเหลือระดับสถานศึกษาของคะแนนการประเมินผลการรู้เรื่องวิทยาศาสตร์ที่มีการควบคุมตัวแปรระดับนักเรียนและสถานศึกษา (U_{0k}) หรือมูลค่าเพิ่มของสถานศึกษาที่ความผันแปรระหว่างสถานศึกษาอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 (χ^2 =653.075, p=.000) เมื่อพิจารณาเมทริกของความแปรปรวนร่วมหรือสหสัมพันธ์ภายในชั้น (intra-class correlation) ของโมเดลเท่ากับ 0.151 หมายความว่า 15.1% เป็นความแปรปรวนของคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์ระหว่างโรงเรียน อีก 84.9% เป็นความแปรปรวนของคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์ระหว่างนักเรียน การประมาณค่าความเที่ยงของสัมประสิทธิ์ตัวแปรสุ่มในระดับที่ 1 มีค่าเท่ากับ 0.599 แสดงว่าค่าเฉลี่ยคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์มีค่าความเที่ยงตามเกณฑ์ประเมิน ของ George และ Mallery (2003) ระดับค่อนข้างต่ำ โดยมีฟังก์ชั่นล็อคไลค์ลิฮูด เท่ากับ -13,355.61

เมื่อคำนวณสัมประสิทธิ์ทำนายหรืออธิบายได้ (R^2) ของโมเดล Undetected DIF-DSF-DTF พบว่า กลุ่มตัวแปรระดับนักเรียนและกลุ่มตัวแปรบริบทระดับสถานศึกษาสามารถอธิบายสัดความแปรปรวนได้ถึง .49507 (49.507%) ส่วนโมเดล Detected DIF-DSF-DTF พบว่า กลุ่มตัวแปรระดับนักเรียนและกลุ่มตัวแปรบริบทระดับสถานศึกษาสามารถอธิบายสัดความแปรปรวนได้ถึง .49946 (49.946%) หมายความว่า ค่าเฉลี่ยรวมคะแนนการประเมินผลการรู้เรื่องวิทยาศาสตร์ที่มีการควบคุมตัวแปรระดับนักเรียนและสถานศึกษาหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันสามารถอธิบายสัดความแปรปรวนได้มากกว่าข้อสอบก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน รายละเอียดตารางที่ 52

ตารางที่ 52 สรุปค่าสัมประสิทธิ์ของตัวแปรระดับนักเรียนและสถานศึกษาทั้ง 4 โมเดล

ตัวแปร	สัมประสิทธิ์ (Coefficient)			
	Undetected DIF-DSF-DTF		Detected DIF-DSF-DTF	
	null model 1	model 1	null model 2	model 2
ค่าเฉลี่ยรวมคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์ทุกสถานศึกษา	492.674**	465.762**	492.246**	469.839**
ระดับนักเรียน				
ตัวแปรดัมมี่ความเป็นเพศหญิงของนักเรียน		1.253		-2.110
ตัวแปรดัมมี่การใช้เวลาเรียนในการเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา		-4.105		-4.104
เศรษฐานะของครอบครัว		13.183**		13.025**
แหล่งทรัพยากรการศึกษาที่บ้าน		10.709**		9.745**
ความมั่งคั่งของครอบครัว		-16.088**		-15.789**
แหล่งทรัพยากรทางไอซีที		5.037		5.551
ระดับสถานศึกษา				
ตัวแปรดัมมี่สถานศึกษาของรัฐ		36.595**		33.688**
ระดับการขาดแคลนครู		3.544		3.670
ร้อยละงบประมาณของรัฐที่จัดสรรให้สถานศึกษา		-0.414**		-0.379**
ทรัพยากรการเรียน		12.413**		11.247**
สัดส่วนนักเรียนต่อครู		-0.198		-0.331
หลักสูตรการวัดและประเมินผล		2.186		1.979
Variance Component				
ระดับที่ 1 (R)	4676.740	4573.915	4627.528	4533.918
ระดับที่ 2 (U_0)	2180.531	874.872	2002.350	807.468
Random level-1 coefficient INTRCPIT1, B_{0j}				
Reliability estimate	0.778	0.614	0.766	0.599
The value of the log-likelihood function	-13,489.95	-13,370.56	-13,471.09	-13,355.61

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 52 สรุปค่าสัมประสิทธิ์ของตัวแปรระดับนักเรียนและสถานศึกษาทั้ง 4 โมเดล

ตัวแปร	สัมประสิทธิ์ (Coefficient)			
	Undetected DIF-DSF-DTF		Detected DIF-DSF-DTF	
	null model 1	model 1	null model 2	model 2
สหสัมพันธ์ภายในชั้น (Intra-class correlation) $(\tau_{00}/(\tau_{00}+\sigma^2)$	0.318	0.161	0.302	0.151
สัมประสิทธิ์การทำนายหรืออธิบายได้ (R^2)		.49507 (49.507%)		.49946 (49.946%)

หมายเหตุ ** p< .01

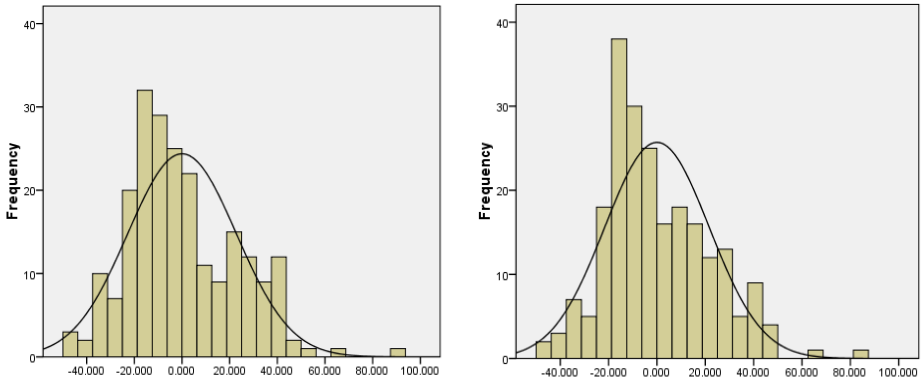
$$R^2 = \frac{(\hat{\tau}_{00}(\text{uncon}) - \hat{\tau}_{00}(\text{hypo})}{\hat{\tau}_{00}(\text{uncon})}$$

4.3 การสร้างเกณฑ์การตัดสินการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม

ผลการวิเคราะห์สถิติพื้นฐานของคะแนนมูลค่าเพิ่มจากการประเมินผลการรู้เรื่องวิทยาศาสตร์จากสถานศึกษาที่มีข้อมูลครบถ้วนตามหลักการวิเคราะห์มูลค่าเพิ่มพหุระดับ จำนวน 223 แห่ง จากสถานศึกษาทั้งหมด 230 แห่ง ของโมเดล Undetected DIF-DSF-DTF และโมเดล Detected DIF-DSF-DTF พบว่า ค่าสถิติของคะแนนมูลค่าเพิ่มพหุระดับของโมเดล Undetected DIF-DSF-DTF และโมเดล Detected DIF-DSF-DTF มีค่าเฉลี่ยเท่ากันคือ 0.00 ซึ่งโมเดล Undetected DIF-DSF-DTF มีค่าสูงสุดเท่ากับ 87.624 และค่าต่ำสุดเท่ากับ -47.648 ช่วงห่างระหว่างค่าสูงสุดและค่าต่ำสุดเท่ากับ 135.272 ส่วนโมเดล Detected DIF-DSF-DTF มีค่าสูงสุดเท่ากับ 82.427 และค่าต่ำสุดเท่ากับ -46.026 ช่วงห่างระหว่างค่าสูงสุดและค่าต่ำสุดเท่ากับ 135.272 ทั้งนี้โมเดล Undetected DIF-DSF-DTF (Sk=0.599, Ku=0.237) และโมเดล Detected DIF-DSF-DTF (Sk=0.623, Ku=0.352) มีคะแนนมูลค่าเพิ่มต่ำกว่าคะแนนมูลค่าเพิ่มเฉลี่ยและการกระจายของคะแนนมูลค่าเพิ่มสูงกว่าโค้งปกติ รายละเอียดตามตารางที่ 53

ตารางที่ 53 ค่าสถิติพื้นฐานของคะแนนมูลค่าเพิ่มพหุระดับจากการประเมินผลการรู้เรื่องวิทยาศาสตร์ของสถานศึกษาทั้ง 2 โมเดล

ค่าสถิติ	โมเดล Undetected DIF-DSF-DTF	โมเดล Detected DIF-DSF-DTF
จำนวนโรงเรียน	223	223
Mean	0.000	0.000
Range	135.272	128.454
Minimum	-47.648	-46.026
Maximum	87.624	82.427
Std. Deviation	22.806	21.636
Variance	520.098	468.109
Skewness	0.599	0.623
Kurtosis	0.237	0.352



รูปภาพที่ 13 ฮีสโตแกรม (Histrogram) คะแนนมูลค่าเพิ่มจากโมเดลการประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์จำนวน 223 แห่ง

เมื่อนำค่าสถิติที่ได้จากการวิเคราะห์คะแนนมูลค่าเพิ่มพหุระดับการประเมินผลการรู้เรื่องวิทยาศาสตร์ของทั้งโมเดล Undetected DIF-DSF-DTF และโมเดล Detected DIF-DSF-DTF ผู้วิจัยได้สร้างเกณฑ์การตัดสินการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม (value-added scores rating: VAS rating) จำนวน 5 ระดับ แบ่งเป็น ระดับ A (ดีมาก) ระดับ B (ดี) ระดับ C (พอใช้) ระดับ D (ผ่าน) และระดับ E (ควรปรับปรุง) ซึ่งระดับคุณภาพที่ได้สามารถให้รายละเอียดที่สะดวกต่อการอธิบายและสามารถนำระดับคุณภาพของทั้งสองโมเดลมาเปรียบเทียบความสอดคล้องของคุณภาพกันได้ รายละเอียดตามตารางที่ 54

ตารางที่ 54 ช่วงคะแนนการตัดสินการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม

ระดับคุณภาพ	คำอธิบาย	ช่วงการตัดสิน	ช่วงคะแนนมูลเพิ่ม	
			Undetected DIF-DSF-DTF Model	Detected DIF-DSF-DTF Model
A	ดีมาก	$\bar{X}$ + 2.5SD ขึ้นไป	57.015 ขึ้นไป	54.09 ขึ้นไป
B	ดี	$\bar{X}$+1.5SD - $\bar{X}$+2.5SD	34.209 - 57.015	32.454 - 54.09
C	พอใช้	$\bar{X}$+0.5SD - $\bar{X}$+1.5SD	11.403 - 34.209	10.818 - 32.454
D	ผ่าน	$\bar{X}$± 0.5SD	(-11.403) - 11.403	(-10.818) - 10.818
E	ควรปรับปรุง	น้อยกว่า $\bar{X}$- 0.5SD	น้อยกว่า (-11.403)	น้อยกว่า (-10.818)

4.4 วิเคราะห์และเปรียบเทียบความสัมพันธ์ระหว่างโมเดลมูลค่าเพิ่มพหุระดับของ 2 โมเดล

ในส่วนนี้เป็นการวิเคราะห์ความสัมพันธ์ระหว่าง 4 ตัวแปร ได้แก่ ตัวแปรคะแนนเฉลี่ยการประเมินผลการรู้เรื่องวิทยาศาสตร์ (SCIEmean) ตัวแปรคะแนนมูลค่าเพิ่ม (VAS) ตัวแปรการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่ม (VAS ranking) และตัวแปรการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม (VAS rating) ก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน (Model 1) และหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน (Model 2) (รายละเอียดคะแนนเฉลี่ยการประเมินผลการรู้เรื่องวิทยาศาสตร์ (SCIEmean) คะแนนมูลค่าเพิ่ม (value-added scores: VAS) การอันดับคุณภาพคะแนนมูลค่าเพิ่ม (value-added scores ranking: VAS ranking) และการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม (value-added scores rating: VAS rating) อยู่ในภาคผนวก ง)

ผลการวิเคราะห์โมเดล Undetected DIF-DSF-DTF เป็นโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ที่ไม่ได้พิจารณาการทำหน้าที่ต่างกันข้อสอบและแบบสอบ พบว่า ตัวแปรคะแนนเฉลี่ยการประเมินผลการรู้เรื่องวิทยาศาสตร์ (SCIEmean) มีความสัมพันธ์กับตัวแปร

คะแนนมูลค่าเพิ่ม (VAS) ตัวแปรจัดการอันดับคุณภาพคะแนนมูลค่าเพิ่ม (VAS ranking) และตัวแปรการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม (VAS rating) อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยที่ตัวแปรคะแนนเฉลี่ยการประเมินผลการรู้เรื่องวิทยาศาสตร์มีความสัมพันธ์กับตัวแปรคะแนนมูลค่าเพิ่มมีความสัมพันธ์ในทิศทางเดียวกัน (r=.737, p<.01) เมื่อหาสัมประสิทธิ์การทำนาย (r^2) ได้เท่ากับ 54.32% หมายความว่า โรงเรียนใดมีคะแนนเฉลี่ยการประเมินผลการรู้เรื่องวิทยาศาสตร์สูงคะแนนมูลค่าเพิ่มก็จะสูง ส่วนโรงเรียนใดมีคะแนนเฉลี่ยการประเมินผลการรู้เรื่องวิทยาศาสตร์ต่ำคะแนนมูลค่าเพิ่มก็จะต่ำ และคะแนนเฉลี่ยการประเมินผลการรู้เรื่องวิทยาศาสตร์อธิบายคะแนนมูลค่าเพิ่มได้เท่ากับ 54.32%

ส่วนความสัมพันธ์ระหว่างตัวแปรคะแนนมูลค่าเพิ่ม (VAS) กับตัวแปรการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่ม (VAS ranking) (r=-.966, p<.01) มีความสัมพันธ์เกือบสมบูรณ์ในทิศทางตรงกันข้าม หาสัมประสิทธิ์การทำนาย (r^2) ได้เท่ากับ 93.32% หมายความว่า โรงเรียนใดมีคะแนนมูลค่าเพิ่มสูงการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่มก็อยู่ในอันดับต้น ส่วนโรงเรียนใดมีคะแนนมูลค่าเพิ่มต่ำการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่มก็อยู่ในอันดับท้าย และคะแนนมูลค่าเพิ่มอธิบายการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่มได้เท่ากับ 93.32% ส่วนความสัมพันธ์ระหว่างตัวแปรคะแนนมูลค่าเพิ่ม (VAS) กับตัวแปรการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม (VAS rating) (r=.946, p<.01) มีความสัมพันธ์เกือบสมบูรณ์ในทิศทางเดียวกัน หาสัมประสิทธิ์การทำนาย (r^2) ได้เท่ากับ 89.49% หมายความว่า โรงเรียนใดมีคะแนนมูลค่าเพิ่มสูงการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่มก็อยู่ในกลุ่มที่มีคุณภาพที่ดี ส่วนโรงเรียนใดมีคะแนนมูลค่าเพิ่มต่ำการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่มก็อยู่ในกลุ่มที่มีคุณภาพที่ไม่ดี และคะแนนมูลค่าเพิ่มอธิบายการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่มได้เท่ากับ 89.49%

ส่วนความสัมพันธ์ระหว่างตัวแปรการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่ม (VAS ranking) และตัวแปรการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม (VAS rating) (r=-.926, p<0.01) มีความสัมพันธ์เกือบสมบูรณ์ในทิศทางตรงกันข้าม หาสัมประสิทธิ์การทำนาย (r^2) ได้เท่ากับ 85.75% หมายความว่า โรงเรียนใดมีการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่มอยู่ในระดับต้นการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่มก็อยู่ในกลุ่มที่มีคุณภาพที่ดี แต่ถ้าโรงเรียนใดมีการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่มอยู่ในระดับท้ายการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่มก็อยู่ในกลุ่มที่มีคุณภาพที่ไม่ดี และคะแนนมูลค่าเพิ่มอธิบายการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่มได้เท่ากับ 85.75%

ผลการวิเคราะห์โมเดล Detected DIF-DSF-DTF เป็นโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ที่พิจารณาการทำหน้าที่ต่างกันข้อสอบและแบบสอบ พบว่า ตัวแปรคะแนนเฉลี่ยการประเมินผลการรู้เรื่องวิทยาศาสตร์ (SCIEmean) มีความสัมพันธ์กับตัวแปรคะแนนมูลค่าเพิ่ม (VAS) ตัวแปรจัดการอันดับคุณภาพคะแนนมูลค่าเพิ่ม (VAS ranking) และตัวแปรการจัด

Chulalongkorn University

กลุ่มคุณภาพคะแนนมูลค่าเพิ่ม (VAS rating) อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยที่ ตัวแปรคะแนนเฉลี่ยการประเมินผลการรู้เรื่องวิทยาศาสตร์มีความสัมพันธ์กับตัวแปรคะแนนมูลค่าเพิ่มมีความสัมพันธ์ในทิศทางเดียวกัน (r=.740, p<.01) เมื่อหาสัมประสิทธิ์การทำนาย (r^2) ได้เท่ากับ 54.76% หมายความว่า โรงเรียนใดมีคะแนนเฉลี่ยการประเมินผลการรู้เรื่องวิทยาศาสตร์สูง คะแนนมูลค่าเพิ่มก็จะสูง ส่วนโรงเรียนใดมีคะแนนเฉลี่ยการประเมินผลการรู้เรื่องวิทยาศาสตร์ต่ำ คะแนนมูลค่าเพิ่มก็จะต่ำ และคะแนนเฉลี่ยการประเมินผลการรู้เรื่องวิทยาศาสตร์อธิบายคะแนนมูลค่าเพิ่มได้เท่ากับ 54.76%

ส่วนความสัมพันธ์ระหว่างตัวแปรคะแนนมูลค่าเพิ่ม (VAS) กับตัวแปรการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่ม (VAS ranking) (r=-.963, p<.01) มีความสัมพันธ์เกือบสมบูรณ์ในทิศทางตรงกันข้ามหาสัมประสิทธิ์การทำนาย (r^2) ได้เท่ากับ 92.74% หมายความว่า โรงเรียนใดมีคะแนนมูลค่าเพิ่มสูงการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่มก็อยู่ในอันดับต้น ส่วนโรงเรียนใดมีคะแนนมูลค่าเพิ่มต่ำการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่มก็อยู่ในอันดับท้าย และคะแนนมูลค่าเพิ่มอธิบายการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่มได้เท่ากับ 92.74% ส่วนความสัมพันธ์ระหว่างตัวแปรคะแนนมูลค่าเพิ่ม (VAS) กับตัวแปรการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม (VAS rating) (r=.945, p<.01) มีความสัมพันธ์เกือบสมบูรณ์ในทิศทางเดียวกัน หาสัมประสิทธิ์การทำนาย (r^2) ได้เท่ากับ 89.30% หมายความว่า โรงเรียนใดมีคะแนนมูลค่าเพิ่มสูงการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่มก็อยู่ในกลุ่มที่มีคุณภาพที่ดี ส่วนโรงเรียนใดมีคะแนนมูลค่าเพิ่มต่ำการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่มก็อยู่ในกลุ่มที่มีคุณภาพที่ไม่ดี และคะแนนมูลค่าเพิ่มอธิบายการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่มได้เท่ากับ 89.30%

ส่วนความสัมพันธ์ระหว่างตัวแปรการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่ม (VAS ranking) และตัวแปรการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม (VAS rating) (r=-.926, p<0.01) มีความสัมพันธ์เกือบสมบูรณ์ในทิศทางตรงกันข้าม หาสัมประสิทธิ์การทำนาย (r^2) ได้เท่ากับ 85.75% หมายความว่า โรงเรียนใดมีการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่มอยู่ในระดับต้นการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่มก็อยู่ในกลุ่มที่มีคุณภาพที่ดี แต่ถ้าโรงเรียนใดมีการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่มอยู่ในระดับท้ายการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่มก็อยู่ในกลุ่มที่มีคุณภาพที่ไม่ดี และคะแนนมูลค่าเพิ่มอธิบายการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่มได้เท่ากับ 85.75% รายละเอียดตามตารางที่ 55

จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 55 ความสัมพันธ์ระหว่าง SCIEmean, VAS, VAS ranking และ VAS rating

ตัวแปร (variables)	โมเดล Undetected DIF-DSF-DTF				โมเดล Detected DIF-DSF-DTF			
	SCIEmean	VAS	VAS ranking	VAS rating	SCIEmean	VAS	VAS ranking	VAS rating
SCIEmean	1.000				1.000			
VAS	.737**	1.000			.740**	1.000		
VAS ranking	-.676**	-.966**	1.000		-.684**	-.963**	1.000	
VAS rating	.714**	.946**	-.926**	1.000	.712**	.945**	-.926**	1.000

หมายเหตุ ** หมายถึง p<.01

ในส่วนของผลการวิเคราะห์ความสัมพันธ์ระหว่างคะแนนความสามารถ (theta scores) และคะแนนมูลค่าเพิ่ม (value-added scores) กับ 3 ตัวแปร ได้แก่ เศรษฐานะของครอบครัว, แหล่งทรัพยากรการศึกษาที่บ้าน, และความมั่งคั่งของครอบครัว ซึ่งเป็นตัวแปรต่อเนื่องระดับนักเรียนที่ถูกนำมาพิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ และการทำหน้าที่ต่างกันเป็นขั้นและเป็นตัวแปรควบคุมในโมเดลวิเคราะห์มูลค่าเพิ่มพหุระดับทั้งโมเดล Undetected DIF-DSF-DTF และโมเดล Detected DIF-DSF-DTF พบว่า ความสัมพันธ์ระหว่างคะแนนค่าความสามารถ (theta scores) กับ 3 ตัวแปร มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และความสัมพันธ์ระหว่างคะแนนมูลค่าเพิ่ม (value-added scores) กับ 3 ตัวแปร มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 เช่นเดียวกัน เมื่อพิจารณาในส่วนคะแนนค่าความสามารถก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า คะแนนค่าความสามารถกับเศรษฐานะของครอบครัวมีความสัมพันธ์สูงสุด (.382) รองลงมาคือ คะแนนค่าความสามารถกับแหล่งทรัพยากรการศึกษาที่บ้าน (.328) นอกจากนี้เมื่อพิจารณาคะแนนค่าความสามารถหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า คะแนนค่าความสามารถกับเศรษฐานะของครอบครัวมีความสัมพันธ์สูงสุด (.377) รองลงมาคือ คะแนนค่าความสามารถกับแหล่งทรัพยากรการศึกษาที่บ้าน (.319)

เมื่อพิจารณาในส่วนคะแนนมูลค่าเพิ่มก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า คะแนนมูลค่าเพิ่มกับเศรษฐานะของครอบครัวมีความสัมพันธ์สูงสุด (.328) รองลงมาคือ คะแนนมูลค่าเพิ่มกับความมั่งคั่งของครอบครัว (.254) นอกจากนี้เมื่อพิจารณาคะแนนมูลค่าเพิ่มหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า คะแนนมูลค่าเพิ่มกับเศรษฐานะของครอบครัวมีความสัมพันธ์สูงสุด (.325) รองลงมาคือ คะแนนมูลค่าเพิ่มกับแหล่งทรัพยากรการศึกษาที่บ้าน (.257)

และเมื่อนำมาเปรียบเทียบความสัมพันธ์ระหว่างคะแนนค่าความสามารถก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันกับ 3 ตัวแปร กับคะแนนมูลค่าเพิ่มหลังตัดข้อสอบที่ตรวจพบการทำ

หน้าที่ต่างกันกับตัวแปร 3 ตัวแปร พบว่า ค่าความสัมพันธ์ระหว่างคะแนนมูลค่าเพิ่มหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันลดลงทั้ง 3 ตัวแปร รายละเอียดตามตารางที่ 56

ตารางที่ 56 ความสัมพันธ์ระหว่างคะแนนจากการประมาณค่าความสามารถและคะแนนมูลค่าเพิ่มระหว่าง 2 โมเดล

ตัวแปร	ความสัมพันธ์			
	$r_{(\text{theta scores – ตัวแปร})}$		$r_{(\text{value-added scores – ตัวแปร})}$	
	ก่อนตัดข้อสอบ DIF	หลังตัดข้อสอบ DIF	ก่อนตัดข้อสอบ DIF	หลังตัดข้อสอบ DIF
เศรษฐานะของครอบครัว	.382**	.377**	.328**	.325**
แหล่งทรัพยากรการศึกษาที่บ้าน	.328**	.319**	.246**	.245**
ความมั่งคั่งของครอบครัว	.273**	.269**	.254**	.257**

หมายเหตุ ** หมายถึง p<.01

4.5 เปรียบเทียบความสอดคล้องการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่มและการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่มของ 2 โมเดล

ผลการวิเคราะห์ความสอดคล้อง ในเบื้องต้นผู้วิจัยนำคะแนนมูลค่าเพิ่มมาพิจาณาตัดสินประสิทธิผลของโรงเรียนสามารถแบ่งโรงเรียนออกเป็น 2 กลุ่ม คือ กลุ่มโรงเรียนที่มีประสิทธิผล (effective school) และกลุ่มโรงเรียนที่ไม่มีประสิทธิผล (ineffective school) โดยที่เกณฑ์การพิจารณาโรงเรียนที่มีคะแนนมูลค่าเพิ่มตั้งแต่ 0 ขึ้นไป เป็นโรงเรียนที่มีประสิทธิผลส่วนโรงเรียนที่มีคะแนนมูลค่าเพิ่มน้อยกว่า 0 หรือติดลบ เป็นโรงเรียนที่ไม่มีประสิทธิผล จากผลการวิเคราะห์พบว่า ทั้งสองโมเดลมีผลการทดสอบประสิทธิผลของโรงเรียนที่สอดคล้องกัน คือ โมเดลการประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ที่ไม่ได้พิจารณาการทำหน้าที่ต่างกันข้อสอบและแบบสอบ พบว่า โรงเรียนที่ไม่มีประสิทธิผลจำนวน 128 โรงเรียน และโรงเรียนที่มีประสิทธิผลจำนวน 95 โรงเรียน ภายหลังจากที่มีการทดสอบความเท่ากันของค่าความแปรปรวน (levene's test) พบว่ามีความแปรปรวนของคะแนนมูลค่าเพิ่มระหว่างกลุ่มโรงเรียนที่มีประสิทธิผลและโรงเรียนที่ไม่มีประสิทธิผลแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 (levene statistic=18.929, p=.00) ส่วนโมเดลการประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ที่พิจารณาข้อสอบที่ทำหน้าที่ต่างกันข้อสอบและการทำหน้าที่ต่างกันของแบบสอบ พบว่า โรงเรียนที่ไม่มีประสิทธิผลจำนวน 128

โรงเรียน และโรงเรียนที่มีประสิทธิผลจำนวน 95 โรงเรียน ภายหลังจากที่มีการทดสอบความเท่ากันของค่าความแปรปรวน (levene's test) พบว่า มีความแปรปรวนของคะแนนมูลค่าเพิ่มระหว่างกลุ่มโรงเรียนที่มีประสิทธิผลและโรงเรียนที่ไม่มีประสิทธิผลแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 (levene statistic=19.133, p=.00) รายละเอียดตามตารางที่ 57

ตารางที่ 57 ผลการเปรียบเทียบความสอดคล้องประสิทธิผลคะแนนมูลค่าเพิ่มของสถานศึกษา

Model		N	ร้อยละ	Levene Statistic	Df1	Df2	p
ก่อนตัดข้อสอบ	Effective school	128	42.6				
	Ineffective school	95	57.4	18.929*	1	221	.00
หลังตัดข้อสอบ	Effective school	95	42.6				
	Ineffective school	128	57.4	19.133*	1	221	.00

หมายเหตุ * หมายถึง p<.05

ผลการเปรียบเทียบการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่ม (VAS ranking) การจัดการศึกษาวิทยาศาสตร์ระหว่างโมเดล Undetected DIF-DSF-DTF กับโมเดล Detected DIF-DSF-DTF ผลการทดสอบความสอดคล้องการจัดอันดับคุณภาพการจัดการศึกษาวิทยาศาสตร์โดยสถิติทดสอบ Wilcoxon Signed Ranks Test พบว่า โมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ตามการจัดอันดับคุณภาพระหว่างโมเดล Undetected DIF-DSF-DTF กับโมเดล Detected DIF-DSF-DTF ไม่มีความสอดคล้องกัน (Z = -.219, p= .829) โดยที่อันดับคุณภาพของ โมเดล Undetected DIF-DSF-DTF มีอันดับน้อยกว่า โมเดล Detected DIF-DSF-DTF จำนวน 97 โรงเรียน ส่วนอันดับคุณภาพของ โมเดล Undetected DIF-DSF-DTF มีอันดับมากกว่าโมเดล Detected DIF-DSF-DTF จำนวน 107 โรงเรียน และอันดับคุณภาพของ โมเดล Undetected DIF-DSF-DTF เท่ากันกับ โมเดล Detected DIF-DSF-DTF จำนวน 19 โรงเรียน รายละเอียดตามตารางที่ 58

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 58 ผลความสอดคล้องการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่มโดยค่าสถิติทดสอบ Wilcoxon Signed Ranks Test

ลักษณะการเปรียบเทียบ	ลักษณะอันดับ	จำนวน	Mean Rank
โมเดล Undetected DIF-DSF-DTF กับโมเดล Detected DIF-DSF-DTF	การจัดอันดับเป็นลบ (-)	97	105.88
	การจัดอันดับเป็นบวก(+)	107	99.43
	การจัดอันดับเท่ากัน (0)	19	
	รวม	**223**	
Wilcoxon Signed Ranks Test; Z = -.219 , p= .829			

หมายเหตุ

(-) หมายถึง อันดับของ โมเดล Undetected DIF-DSF-DTF < โมเดล Detected DIF-DSF-DTF

(+) หมายถึง อันดับของ โมเดล Undetected DIF-DSF-DTF > โมเดล Detected DIF-DSF-DTF

(0) หมายถึง อันดับของ โมเดล Undetected DIF-DSF-DTF = โมเดล Detected DIF-DSF-DTF

ส่วนผลการทดสอบความสอดคล้องของการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม (VAS rating) การจัดการศึกษาวิทยาศาสตร์ระหว่างโมเดล Undetected DIF-DSF-DTF กับโมเดล Detected DIF-DSF-DTF ตามกลุ่มคุณภาพการจัดการศึกษา 5 ระดับ คือ ระดับ A (ดีมาก) ระดับ B (ดี) ระดับ C (พอใช้) ระดับ D (ผ่าน) และระดับ E (ควรปรับปรุง) ผลการทดสอบความสอดคล้องการจัดระดับคุณภาพการจัดการศึกษาโดยอาศัยหลักเกณฑ์การวัดความสอดคล้อง (Measure of Agreement) ของ Cohen's Kappa ซึ่งใช้ค่าสัมประสิทธิ์ของ Kappa เป็นตัวชี้วัดทางสถิติระหว่างผู้ให้ความคิดเห็นสองฝ่ายว่ามีความเห็นสอดคล้องกันมากหรือน้อยอย่างไร พบว่า โมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ตามการจัดกลุ่มคุณภาพระหว่างโมเดล Undetected DIF-DSF-DTF กับโมเดล Detected DIF-DSF-DTF มีผลการจัดกลุ่มคุณภาพโรงเรียนสอดคล้องกันระดับค่อนข้างสมบูรณ์ (Kappa = .865) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

โดยที่สถานศึกษาตามโมเดล Undetected DIF-DSF-DTF ส่วนใหญ่มีระดับคุณภาพ E (ควรปรับปรุง) มากที่สุด (81 โรงเรียน) รองลงมาคือ ระดับคุณภาพ D (ผ่าน) C (พอใช้) B (ดี) และ A (ดีมาก) ตามลำดับ (80, 41, 19, และ 2 โรงเรียน) ส่วนโมเดล Detected DIF-DSF-DTF ส่วนใหญ่มีระดับคุณภาพ E (ควรปรับปรุง) มากที่สุด (83 โรงเรียน) รองลงมาคือ ระดับคุณภาพ D (ผ่าน) C (พอใช้) B (ดี) และ A (ดีมาก) ตามลำดับ (78, 42, 18, และ 2 โรงเรียน) ทั้งนี้ผลการทดสอบความสอดสอดคล้องการจัดกลุ่มคุณภาพของสถานศึกษา พบว่า โมเดล Undetected DIF-DSF-DTF สอดคล้องกับโมเดล Detected DIF-DSF-DTF โดยภาพรวมจำนวน 202 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 90.58 ซึ่งที่ระดับคุณภาพ E (ควรปรับปรุง) มีจำนวนสถานศึกษา

สอดคล้องกัน 74 โรงเรียน ระดับคุณภาพ D (ผ่าน) มีจำนวนสถานศึกษาสอดคล้องกัน 70 โรงเรียน ระดับคุณภาพ C (พอใช้) มีจำนวนสถานศึกษาสอดคล้องกัน 39 โรงเรียน ระดับคุณภาพ B (ดี) มีจำนวนสถานศึกษาสอดคล้องกัน 17 โรงเรียน และระดับคุณภาพ A (ดีมาก) มีจำนวนสถานศึกษาสอดคล้องกัน 2 โรงเรียน รายละเอียดตามตารางที่ 59

ตารางที่ 59 ผลความสอดคล้องการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม (VAS rating) โดยค่าสัมประสิทธิ์ของ Kappa

ระดับคุณภาพ		โมเดล Detected DIF-DSF-DTF					
		E (ควรรับปรุง)	D (ผ่าน)	C (พอใช้)	B (ดี)	A (ดีมาก)	รวม
โมเดล Undetected DIF-DSF-DTF	E (ควรปรับปรุง)	*74*	7	0	0	0	**81**
	D (ผ่าน)	9	*70*	1	0	0	**80**
	C (พอใช้)	0	1	*39*	1	0	**41**
	B (ดี)	0	0	2	*17*	0	**19**
	A (ดีมาก)	0	0	0	0	*2*	**2**
	รวม	**83**	**78**	**42**	**18**	**2**	**223**
Cohen's Kappa = .865*, p = .000							

หมายเหตุ * หมายถึง $p<.05$

ในส่วนผลการทดสอบความสอดคล้องการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม (VAS rating) 5 ระดับ ระหว่างโมเดล Undetected DIF-DSF-DTF กับโมเดล Detected DIF-DSF-DTF ตามตัวแปรสถานศึกษา ได้แก่ ขนาดสถานศึกษา ประเภทสถานศึกษา สังกัดของสถานศึกษา ตำแหน่งที่ตั้งของสถานศึกษา และระดับการขาดแคลนครูวิทยาศาสตร์ พบว่า

ขนาดสถานศึกษา จากการทดสอบความสอดคล้องการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่มระหว่างโมเดล Undetected DIF-DSF-DTF กับโมเดล Detected DIF-DSF-DTF เมื่อแยกพิจารณาเป็น**ขนาดสถานศึกษาขนาดเล็ก** พบว่า จากจำนวนสถานศึกษา 169 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 152 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 89.94 แยกเป็นระดับ E (ควรปรับปรุง) จำนวน 55 โรงเรียน, ระดับ D (ผ่าน) จำนวน 58 โรงเรียน, ระดับ C (พอใช้) จำนวน 28 โรงเรียน, ระดับ B (ดี) จำนวน 10 โรงเรียน, และระดับ A (ดีมาก) จำนวน 1 โรงเรียน **ขนาดสถานศึกษาขนาดกลาง** พบว่า จากจำนวนสถานศึกษา 51 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 47 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 92.16 แยกเป็นระดับ E (ควรปรับปรุง) จำนวน

18 โรงเรียน, ระดับ D (ผ่าน) จำนวน 11 โรงเรียน, ระดับ C (พอใช้) จำนวน 11 โรงเรียน, และระดับ B (ดี) จำนวน 7 โรงเรียน, และ**ขนาดสถานศึกษาขนาดใหญ่** พบว่า จากจำนวนสถานศึกษา 3 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 3 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 100 แยกเป็นระดับ E (ควรปรับปรุง) จำนวน 1 โรงเรียน, ระดับ D (ผ่าน) จำนวน 1 โรงเรียน, และระดับ A (ดีมาก) จำนวน 1 โรงเรียน

ประเภทสถานศึกษา จากการทดสอบความสอดคล้องการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่มระหว่างโมเดล Undetected DIF-DSF-DTF กับโมเดล Detected DIF-DSF-DTF เมื่อแยกพิจารณาเป็น**ประเภทสถานศึกษาของรัฐ** พบว่า จากจำนวนสถานศึกษา 197 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 178 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 90.36 แยกเป็นระดับ E (ควรปรับปรุง) จำนวน 64 โรงเรียน, ระดับ D (ผ่าน) จำนวน 65 โรงเรียน, ระดับ C (พอใช้) จำนวน 34 โรงเรียน, ระดับ B (ดี) จำนวน 13 โรงเรียน, และระดับ A (ดีมาก) จำนวน 2 โรงเรียน และ**ประเภทสถานศึกษาเอกชน** พบว่า จากจำนวนสถานศึกษา 26 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 24 โรงเรียน คิดเป็น ร้อยละที่สอดคล้องร้อยละ 92.31 แยกเป็นระดับ E (ควรปรับปรุง) จำนวน 10 โรงเรียน, ระดับ D (ผ่าน) จำนวน 5 โรงเรียน, ระดับ C (พอใช้) จำนวน 5 โรงเรียน, และระดับ B (ดี) จำนวน 4 โรงเรียน

สังกัดของสถานศึกษา จากการทดสอบความสอดคล้องการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม ระหว่างโมเดล Undetected DIF-DSF-DTF กับโมเดล Detected DIF-DSF-DTF เมื่อแยกพิจารณาเป็น**โรงเรียนในสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานที่มาจากโรงเรียนขยายโอกาสเดิม (สพฐ.1)** พบว่า จากจำนวนสถานศึกษา 35 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 30 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 85.71 แยกเป็นระดับ E (ควรปรับปรุง) จำนวน 11 โรงเรียน, ระดับ D (ผ่าน) จำนวน 16 โรงเรียน, และระดับ C (พอใช้) จำนวน 3 โรงเรียน **โรงเรียนในสังกัดกรมสามัญศึกษาเดิม (สพฐ.2)** พบว่า จากจำนวนสถานศึกษา 91 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 82 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 90.11 แยกเป็นระดับ E (ควรปรับปรุง) จำนวน 26 โรงเรียน, ระดับ D (ผ่าน) จำนวน 25 โรงเรียน, ระดับ C (พอใช้) จำนวน 21 โรงเรียน, ระดับ B (ดี) จำนวน 9 โรงเรียน, และระดับ A (ดีมาก) จำนวน 1 โรงเรียน **โรงเรียนในสังกัดสำนักบริหารงานคณะกรรมการส่งเสริมการศึกษาเอกชน (สช.)** พบว่า จากจำนวนสถานศึกษา 13 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 13 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 100.00 แยกเป็นระดับ E (ควรปรับปรุง) จำนวน 4 โรงเรียน, ระดับ D (ผ่าน) จำนวน 3 โรงเรียน, ระดับ C (พอใช้) จำนวน 3 โรงเรียน, และระดับ B (ดี) จำนวน 3 โรงเรียน **โรงเรียนในสังกัดสำนักการศึกษากรุงเทพมหานคร (กทม.)** พบว่า จากจำนวนสถานศึกษา 20 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 19 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 95.00 แยก

เป็นระดับ E (ควรปรับปรุง) จำนวน 10 โรงเรียน, ระดับ D (ผ่าน) จำนวน 8 โรงเรียน, และระดับ B (ดี) จำนวน 1 โรงเรียน **โรงเรียนในสังกัดสำนักประสานและพัฒนาการจัดการศึกษาท้องถิ่น (กศท.)** พบว่า จากจำนวนสถานศึกษา 17 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 16 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 94.12 แยกเป็นเป็นระดับ E (ควรปรับปรุง) จำนวน 8 โรงเรียน, ระดับ D (ผ่าน) จำนวน 6 โรงเรียน, ระดับ C (พอใช้) จำนวน 1 โรงเรียน, และระดับ B (ดี) จำนวน 1 โรงเรียน **โรงเรียนในสังกัดสาธิตของมหาวิทยาลัย (สาธิต)** พบว่า จากจำนวนสถานศึกษา 16 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 15 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 93.75 แยกเป็นระดับ E (ควรปรับปรุง) จำนวน 2 โรงเรียน, ระดับ D (ผ่าน) จำนวน 3 โรงเรียน, ระดับ C (พอใช้) จำนวน 7 โรงเรียน, ระดับ B (ดี) จำนวน 2 โรงเรียน, และระดับ A (ดีมาก) จำนวน 1 โรงเรียน **โรงเรียนอาชีวศึกษาของเอกชน (อศ.1)** พบว่า จากจำนวนสถานศึกษา 13 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 11 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 84.62 แยกเป็นระดับ E (ควรปรับปรุง) จำนวน 6 โรงเรียน, ระดับ D (ผ่าน) จำนวน 2 โรงเรียน, ระดับ C (พอใช้) จำนวน 2 โรงเรียน, และระดับ B (ดี) จำนวน 1 โรงเรียน และ **วิทยาลัยอาชีวศึกษาของรัฐ (อศ.2)** พบว่า จากจำนวนสถานศึกษา 18 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 16 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 88.89 แยกเป็นระดับ E (ควรปรับปรุง) จำนวน 7 โรงเรียน, ระดับ D (ผ่าน) จำนวน 7 โรงเรียน, และ ระดับ C (พอใช้) 2 โรงเรียน

ตำแหน่งที่ตั้งของสถานศึกษา จากการทดสอบความสอดคล้องการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม ระหว่างโมเดล Undetected DIF-DSF-DTF กับโมเดล Detected DIF-DSF-DTF เมื่อแยกพิจารณาเป็น**สถานศึกษาที่ตั้งอยู่ในหมู่บ้าน** พบว่า จากจำนวนสถานศึกษา 47 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 40 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 85.11 แยกเป็นระดับ E (ควรปรับปรุง) จำนวน 14 โรงเรียน, ระดับ D (ผ่าน) จำนวน 20 โรงเรียน, และระดับ C (พอใช้) จำนวน 6 โรงเรียน, **สถานศึกษาที่ตั้งอยู่ในเมืองเล็ก** พบว่า จากจำนวนสถานศึกษา 45 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 43 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 95.56 แยกเป็นระดับ E (ควรปรับปรุง) จำนวน 21 โรงเรียน, ระดับ D (ผ่าน) จำนวน 17 โรงเรียน, ระดับ C (พอใช้) จำนวน 4 โรงเรียน, และระดับ B (ดี) จำนวน 1 โรงเรียน **สถานศึกษาที่ตั้งอยู่ในเมือง** พบว่า จากจำนวนสถานศึกษา 73 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 65 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 89.04 แยกเป็นระดับ E (ควรปรับปรุง) จำนวน 25 โรงเรียน, ระดับ D (ผ่าน) จำนวน 20 โรงเรียน, ระดับ C (พอใช้) จำนวน 11 โรงเรียน, และระดับ B (ดี) จำนวน 9 โรงเรียน **สถานศึกษาที่ตั้งอยู่ในเมืองใหญ่** พบว่า จากจำนวนสถานศึกษา 38 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 34 โรงเรียน คิดเป็นร้อยละที่สอดคล้อง ร้อยละ 89.47 แยกเป็นระดับ E

(ควรปรับปรุง) จำนวน 9 โรงเรียน, ระดับ D (ผ่าน) จำนวน 9 โรงเรียน, ระดับ C (พอใช้) จำนวน 13 โรงเรียน, และระดับ B (ดี) จำนวน 3 โรงเรียน และ**สถานศึกษาที่ตั้งอยู่ในเมืองใหญ่มาก** พบว่า จากจำนวนสถานศึกษา 20 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 20 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 100.00 แยกเป็นระดับ E (ควรปรับปรุง) จำนวน 5 โรงเรียน, ระดับ D (ผ่าน) จำนวน 4 โรงเรียน, ระดับ C (พอใช้) จำนวน 5 โรงเรียน, ระดับ B (ดี) จำนวน 4 โรงเรียน, และระดับ A (ดีมาก) จำนวน 2 โรงเรียน

ระดับการขาดแคลนครูวิทยาศาสตร์ จากการทดสอบความสอดคล้องการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม ระหว่างโมเดล Undetected DIF-DSF-DTF กับโมเดล Detected DIF-DSF-DTF เมื่อแยกพิจารณาเป็น**ระดับไม่ขาดแคลนครูวิทยาศาสตร์** พบว่า จากจำนวนสถานศึกษา 86 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 77 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 89.53 แยกเป็นระดับ E (ควรปรับปรุง) จำนวน 31 โรงเรียน, ระดับ D (ผ่าน) จำนวน 29 โรงเรียน, ระดับ C (พอใช้) จำนวน 10 โรงเรียน, ระดับ B (ดี) จำนวน 6 โรงเรียน, และระดับ A (ดีมาก) จำนวน 1 โรงเรียน **ระดับขาดแคลนครูวิทยาศาสตร์น้อย** พบว่า จากจำนวนสถานศึกษา 41 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 40 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 97.56 แยกเป็นระดับ E (ควรปรับปรุง) จำนวน 13 โรงเรียน, ระดับ D (ผ่าน) จำนวน 13 โรงเรียน, ระดับ C (พอใช้) จำนวน 10 โรงเรียน, และระดับ B (ดี) จำนวน 4 โรงเรียน **ระดับขาดแคลนครูวิทยาศาสตร์เป็นบางส่วน** พบว่า จากจำนวนสถานศึกษา 81 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 72 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 88.89 แยกเป็นระดับ E (ควรปรับปรุง) จำนวน 24 โรงเรียน, ระดับ D (ผ่าน) จำนวน 24 โรงเรียน, ระดับ C (พอใช้) จำนวน 16 โรงเรียน, ระดับ B (ดี) จำนวน 7 โรงเรียน, และระดับ A (ดีมาก) จำนวน 1 โรงเรียน และ**ระดับขาดแคลนครูวิทยาศาสตร์จำนวนมาก** พบว่า จากจำนวนสถานศึกษา 15 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 13 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 86.67 แยกเป็นระดับ E (ควรปรับปรุง) จำนวน 6 โรงเรียน, ระดับ D (ผ่าน) จำนวน 4 โรงเรียน, และระดับ C (พอใช้) จำนวน 3 โรงเรียน รายละเอียดตามรางที่ 60

ตารางที่ 60 เปรียบเทียบความสอดคล้องตามตัวแปรสถานศึกษา

ระดับคุณภาพ			Model 2					
			E (ควรรับปรุง)	D (ผ่าน)	C (พอใช้)	B (ดี)	A (ดีมาก)	รวม
1.ขนาดสถานศึกษา								
Model 1	1.1 เล็ก	E (ควรปรับปรุง)	*55*	7	0	0	0	62
		D (ผ่าน)	8	*58*	0	0	0	66
		C (พอใช้)	0	0	*28*	1	0	29
		B (ดี)	0	0	1	*10*	0	11
		A (ดีมาก)	0	0	0	0	*1*	1
		รวม	63	65	29	11	1	169
	1.2 กลาง	E (ควรปรับปรุง)	*18*	0	0	0	0	18
		D (ผ่าน)	1	*11*	1	0	0	13
		C (พอใช้)	0	1	*11*	0	0	12
		B (ดี)	0	0	1	*7*	0	8
		A (ดีมาก)	0	0	0	0	0	0
		รวม	19	12	13	7	0	51
	1.1 ใหญ่	E (ควรปรับปรุง)	*1*	0	0	0	0	1
		D (ผ่าน)	0	*1*	0	0	0	1
		C (พอใช้)	0	0	0	0	0	0
		B (ดี)	0	0	0	0	0	0
		A (ดีมาก)	0	0	0	0	*1*	1
		รวม	1	1	0	0	1	3
2. ประเภทสถานศึกษา								
Model 1	2.1 รัฐ	E (ควรปรับปรุง)	*64*	7	0	0	0	71
		D (ผ่าน)	7	65	1	0	0	73
		C (พอใช้)	0	1	*34*	1	0	36
		B (ดี)	0	0	2	*13*	0	15
		A (ดีมาก)	0	0	0	0	*2*	2
		รวม	71	73	37	14	2	197

ตารางที่ 60 เปรียบเทียบความสอดคล้องตามตัวแปรสถานศึกษา(ต่อ)

ระดับคุณภาพ			Model 2					
			E (ควรรับปรุง)	D (ผ่าน)	C (พอใช้)	B (ดี)	A (ดีมาก)	รวม
2. ประเภทสถานศึกษา (ต่อ)								
	2.2 เอกชน	**E (ควรปรับปรุง)**	*10*	0	0	0	0	**10**
		D (ผ่าน)	2	*5*	0	0	0	**7**
		C (พอใช้)	0	0	*5*	0	0	**5**
		B (ดี)	0	0	0	*4*	0	**4**
		A (ดีมาก)	0	0	0	0	0	**0**
		รวม	**12**	**5**	**5**	**4**	**0**	**26**
3. สังกัดของสถานศึกษา								
Model 1	**3.1** สพฐ. 1	**E (ควรปรับปรุง)**	*11*	1	0	0	0	**12**
		D (ผ่าน)	3	*16*	0	0	0	**19**
		C (พอใช้)	0	0	*3*	0	0	**3**
		B (ดี)	0	0	1	0	0	**1**
		A (ดีมาก)	0	0	0	0	0	**0**
		รวม	**14**	**17**	**4**	**0**	**0**	**35**
	3.2 สพฐ. 2	**E (ควรปรับปรุง)**	*26*	2	0	0	0	**28**
		D (ผ่าน)	4	*25*	1	0	0	**30**
		C (พอใช้)	0	0	*21*	1	0	**22**
		B (ดี)	0	0	1	*9*	0	**10**
		A (ดีมาก)	0	0	0	0	*1*	**1**
		รวม	**30**	**27**	**23**	**10**	**1**	**91**
	3.3 สช.	**E (ควรปรับปรุง)**	*4*	0	0	0	0	**4**
		D (ผ่าน)	0	*3*	0	0	0	**3**
		C (พอใช้)	0	0	*3*	0	0	**3**
		B (ดี)	0	0	0	*3*	0	**3**
		A (ดีมาก)	0	0	0	0	0	**0**
		รวม	**4**	**3**	**3**	**3**	**0**	**13**

ตารางที่ 60 เปรียบเทียบความสอดคล้องตามตัวแปรสถานศึกษา(ต่อ)

ระดับคุณภาพ			Model 2					
			E (ควรรับปรุง)	D (ผ่าน)	C (พอใช้)	B (ดี)	A (ดีมาก)	รวม
3. สังกัดของสถานศึกษา(ต่อ)								
	3.4 กทม.	E (ควรปรับปรุง)	*10*	1	0	0	0	11
		D (ผ่าน)	0	*8*	0	0	0	8
		C (พอใช้)	0	0	0	0	0	0
		B (ดี)	0	0	0	*1*	0	1
		A (ดีมาก)	0	0	0	0	0	0
		รวม	10	9	0	1	0	20
	3.5 กศท.	E (ควรปรับปรุง)	*8*	1	0	0	0	9
		D (ผ่าน)	0	*6*	0	0	0	6
		C (พอใช้)	0	0	*1*	0	0	1
		B (ดี)	0	0	0	*1*	0	1
		A (ดีมาก)	0	0	0	0	0	0
		รวม	8	7	1	1	0	17
	3.6 สาธิต	E (ควรปรับปรุง)	*2*	1	0	0	0	3
		D (ผ่าน)	0	*3*	0	0	0	3
		C (พอใช้)	0	0	*7*	0	0	7
		B (ดี)	0	0	0	*2*	0	2
		A (ดีมาก)	0	0	0	0	*1*	1
		รวม	2	4	7	2	1	16
	3.7 อศ.1	E (ควรปรับปรุง)	*6*	0	0	0	0	6
		D (ผ่าน)	2	*2*	0	0	0	4
		C (พอใช้)	0	0	*2*	0	0	2
		B (ดี)	0	0	0	*1*	0	1
		A (ดีมาก)	0	0	0	0	0	0
		รวม	8	2	2	1	0	13

ตารางที่ 60 เปรียบเทียบความสอดคล้องตามตัวแปรสถานศึกษา(ต่อ)

ระดับคุณภาพ			Model 2					
			E (ควรรับปรุง)	D (ผ่าน)	C (พอใช้)	B (ดี)	A (ดีมาก)	รวม
3. สังกัดของสถานศึกษา(ต่อ)								
	3.8 อศ.2	**E (ควรปรับปรุง)**	*7*	1	0	0	0	**8**
		D (ผ่าน)	0	*7*	0	0	0	**7**
		C (พอใช้)	0	1	*2*	0	0	**3**
		B (ดี)	0	0	0	0	0	**0**
		A (ดีมาก)	0	0	0	0	0	**0**
		รวม	**7**	**9**	**2**	**0**	**0**	**18**
4. ตำแหน่งที่ตั้งของสถานศึกษา								
Model 1	**4.1 หมู่บ้าน**	**E (ควรปรับปรุง)**	*14*	3	0	0	0	**17**
		D (ผ่าน)	3	*20*	0	0	0	**23**
		C (พอใช้)	0	0	*6*	1	0	**7**
		B (ดี)	0	0	0	0	0	**0**
		A (ดีมาก)	0	0	0	0	0	**0**
		รวม	**17**	**23**	**6**	**1**	**0**	**47**
	4.2 เมืองเล็ก	**E (ควรปรับปรุง)**	*21*	0	0	0	0	**21**
		D (ผ่าน)	2	*17*	0	0	0	**19**
		C (พอใช้)	0	0	*4*	0	0	**4**
		B (ดี)	0	0	0	*1*	0	**1**
		A (ดีมาก)	0	0	0	0	0	**0**
		รวม	**23**	**17**	**4**	**1**	**0**	**45**
	4.3 เมือง	**E (ควรปรับปรุง)**	*25*	2	0	0	0	**27**
		D (ผ่าน)	3	*20*	0	0	0	**23**
		C (พอใช้)	0	1	*11*	0	0	**12**
		B (ดี)	0	0	2	*9*	0	**11**
		A (ดีมาก)	0	0	0	0	*0*	**0**
		รวม	**28**	**23**	**13**	**9**	**0**	**73**

ตารางที่ 60 เปรียบเทียบความสอดคล้องตามตัวแปรสถานศึกษา(ต่อ)

ระดับคุณภาพ			Model 2					
			E (ควรรับปรุง)	D (ผ่าน)	C (พอใช้)	B (ดี)	A (ดีมาก)	รวม
4. ตำแหน่งที่ตั้งของสถานศึกษา (ต่อ)								
	4.4 เมืองใหญ่	E (ควรปรับปรุง)	*9*	2	0	0	0	11
		D (ผ่าน)	1	*9*	1	0	0	11
		C (พอใช้)	0	0	*13*	0	0	13
		B (ดี)	0	0	0	*3*	0	3
		A (ดีมาก)	0	0	0	0	*0*	0
		รวม	10	11	14	3	0	38
	4.5 เมืองใหญ่มาก	E (ควรปรับปรุง)	*5*	0	0	0	0	5
		D (ผ่าน)	0	*4*	0	0	0	4
		C (พอใช้)	0	0	*5*	0	0	5
		B (ดี)	0	0	0	*4*	0	4
		A (ดีมาก)	0	0	0	0	*2*	2
		รวม	5	4	5	4	2	20
5. ระดับการขาดแคลนครูวิทยาศาสตร์								
Model 1	5.1 ไม่ขาดแคลน	E (ควรปรับปรุง)	*31*	4	0	0	0	35
		D (ผ่าน)	4	*29*	1	0	0	34
		C (พอใช้)	0	0	*10*	0	0	10
		B (ดี)	0	0	0	*6*	0	6
		A (ดีมาก)	0	0	0	0	1	1
		รวม	35	33	11	6	1	86
	5.2 ขาดแคลนน้อย	E (ควรปรับปรุง)	*13*	1	0	0	0	14
		D (ผ่าน)	0	*13*	0	0	0	13
		C (พอใช้)	0	0	*10*	0	0	10
		B (ดี)	0	0	0	*4*	0	4
		A (ดีมาก)	0	0	0	0	0	0
		รวม	13	14	10	4	0	41

ตารางที่ 60 เปรียบเทียบความสอดคล้องตามตัวแปรสถานศึกษา(ต่อ)

ระดับคุณภาพ			Model 2					
			E (ควรรับปรุง)	D (ผ่าน)	C (พอใช้)	B (ดี)	A (ดีมาก)	รวม
5. ระดับการขาดแคลนครูวิทยาศาสตร์(ต่อ)								
	5.3 ขาดแคลนเป็นบางส่วน	E (ควรปรับปรุง)	*24*	1	0	0	0	25
		D (ผ่าน)	5	*24*	0	0	0	29
		C (พอใช้)	0	1	*16*	1	0	18
		B (ดี)	0	0	1	*7*	0	8
		A (ดีมาก)	0	0	0	0	*1*	1
		รวม	*29*	26	17	8	1	81
	5.4 ขาดแคลนจำนวนมาก	E (ควรปรับปรุง)	*6*	1	0	0	0	7
		D (ผ่าน)	0	*4*	0	0	0	4
		C (พอใช้)	0	0	*3*	0	0	3
		B (ดี)	0	0	1	0	0	1
		A (ดีมาก)	0	0	0	0	0	0
		รวม	*6*	5	4	0	0	15

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 5
สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

ในการวิจัยครั้งนี้เป็นการศึกษาผลการทำหน้าที่ต่างกันของข้อสอบ (DIF) ที่มีการตรวจให้คะแนนทวิวิภาคและพหุวิภาค และการทำหน้าที่ต่างกันของแบบทดสอบ (DTF) ในแบบทดสอบการประเมินการรู้เรื่องวิทยาศาสตร์ และเปรียบเทียบประสิทธิภาพของโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์เมื่อใช้โมเดลมูลค่าเพิ่มพหุระดับ 2 โมเดล คือ โมเดลที่ไม่ได้พิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ (โมเดลที่ 1) กับโมเดลที่มีการพิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบโดยการตัดข้อสอบที่มีการทำหน้าที่ต่างกันของข้อสอบและการทำหน้าที่ต่างกันเป็นขั้นที่แตกต่างกัน (โมเดลที่ 2) ซึ่งในการเปรียบเทียบประสิทธิภาพของโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ มีวัตถุประสงค์ย่อยคือ เพื่อศึกษาความสัมพันธ์ระหว่างคะแนนมูลค่าเพิ่มกับ 3 ตัวแปร คือ เศรษฐานะของครอบครัว แหล่งทรัพยากรการศึกษาที่บ้าน และความมั่งคั่งของครอบครัว ทั้งก่อนและหลังตัดข้อสอบที่ทำหน้าที่ต่างกัน และเพื่อเปรียบเทียบความสอดคล้องของการจัดอันดับคุณภาพและการจัดกลุ่มคุณภาพของโมเดลมูลค่าเพิ่มพหุระดับทั้ง 2 โมเดล

จากการประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ของสถานศึกษาตามการศึกษาขั้นพื้นฐานประกอบด้วยโมเดลที่ใช้ในการวิจัย 2 โมเดล คือ โมเดล Undetected DIF-DSF-DTF ซึ่งเป็นโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ที่ไม่ได้พิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ ส่วนโมเดลการวิเคราะห์มูลค่าเพิ่มพหุระดับเป็นโมเดลที่มีควบคุมตัวแปรระดับนักเรียนและสถานศึกษา ส่วนโมเดล Detected DIF-DSF-DTF ซึ่งเป็นโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ที่พิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ โดยการตัดข้อสอบภายหลังจากที่มีการพิจารณาการทำหน้าที่ต่างกันของข้อสอบ การทำหน้าที่ต่างกันเป็นขั้น ส่วนโมเดลการวิเคราะห์มูลค่าเพิ่มพหุระดับเป็นโมเดลที่มีการควบคุมตัวแปรระดับนักเรียนและสถานศึกษา

ทั้งนี้แบบสอบที่มีการตรวจให้คะแนนทวิวิภาคใช้การตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันด้วย วิธีการวิเคราะห์แมนเทล-แฮนส์เซลร่วมกับอัตราส่วนแต้มต่อมาตรฐาน (standardized mantel-haenszel log-odds ratio: LOR Z) สำหรับข้อสอบที่มีการให้คะแนนพหุวิภาคใช้การตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันโดย สถิติตรวจสอบทำหน้าที่ต่างกันเป็นขั้นด้วยตัวประมาณค่าอิทธิพลดีเอสเอฟกับอัตราส่วนแต้มต่อระดับขั้น (a step-level log-odds ratio DSF effect estimator) แบบแนวคิดสะสม (the cumulative approach: CU-LOR) ร่วมกับตัวประมาณค่า

อิทธิพลการทำหน้าที่ต่างกันของข้อสอบอัตราส่วนแต้มต่อลัว-อเกรสตีปกติที่เป็นมาตรฐาน (standardized Liu-Agresti common log-odds ratio DIF effect estimator: Z(Lor)) สำหรับการศึกษาการทำหน้าที่ต่างกันของแบบสอบในแบบสอบที่มีการตรวจให้คะแนนแบบผสมผสาน (mixed format tests) ใช้การตรวจสอบความแปรปรวนอิทธิพลการทำหน้าที่ต่างกันของข้อสอบ (the DIF effect variance)

5.1 สรุปผลการวิจัย

จากการวิเคราะห์และเปรียบเทียบโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์โดยการประยุกต์ใช้โมเดลมูลค่าเพิ่มพหุระดับที่มีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ ผู้วิจัยสรุปผลการวิจัยออกเป็น 3 ประเด็นหลักดังต่อไปนี้

ประเด็นที่ 1 ผลจากการวิเคราะห์ข้อมูลพื้นฐานในการวิจัย

ผลการวิเคราะห์ข้อมูลพื้นฐานของกลุ่มตัวอย่างตามโครงการประเมินผลนักเรียนนานาชาติ (PISA) ตามตัวแปรระดับนักเรียนและสถานศึกษา สำหรับตัวแปรจัดประเภท จำแนกตามคุณลักษณะของนักเรียนและสถานศึกษา ในระดับนักเรียน พบว่า นักเรียนที่ทำการศึกษาส่วนใหญ่เป็นนักเรียนหญิง (ร้อยละ 56.8) และนักเรียนส่วนใหญ่ไม่มีเวลาเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา (ร้อยละ 25.3) ส่วนในระดับสถานศึกษา พบว่า สถานศึกษาส่วนใหญ่มีขนาดเล็ก (ร้อยละ 76.5) ซึ่งเป็นสถานศึกษาของรัฐ (ร้อยละ 88.7) ส่วนสังกัดสถานศึกษา พบว่า นักเรียนส่วนใหญ่มาจากโรงเรียนในสังกัดกรมสามัญศึกษาเดิม (สพฐ.2) (ร้อยละ 39.6) ตำแหน่งที่ตั้งของสถานศึกษา พบว่า ส่วนใหญ่ตั้งอยู่ในเมือง (ร้อยละ 31.7) และเมื่อพิจารณาระดับการขาดแคลนครูวิทยาศาสตร์ในโรงเรียน พบว่า โรงเรียนส่วนใหญ่ไม่ขาดแคลนครูวิทยาศาสตร์ (ร้อยละ 39.1)

สำหรับตัวแปรต่อเนื่องในระดับนักเรียน พบว่า เศรษฐานะของครอบครัวมีค่าเฉลี่ยเท่ากับ -1.16 แหล่งทรัพยากรที่บ้านมีค่าเฉลี่ยเท่ากับ -0.23 ความมั่งคั่งของครอบครัวมีค่าเฉลี่ยเท่ากับ -1.16 และแหล่งทรัพยากรทางไอซีทีมีค่าเฉลี่ยเท่ากับ -.83 โดยที่เศรษฐานะของครอบครัว แหล่งทรัพยากรการศึกษาที่บ้าน ความมั่งคั่งของครอบครัว และแหล่งทรัพยากรทางไอซีที มีการกระจายข้อมูลน้อยกว่าค่าเฉลี่ยและต่ำกว่าโค้งปกติ ส่วนระดับสถานศึกษา พบว่า ที่ร้อยละงบประมาณของรัฐที่จัดสรรให้มีค่าเฉลี่ยเท่ากับ 81.91 ทรัพยากรการเรียนมีค่าเฉลี่ยเท่ากับ -0.41 สัดส่วนนักเรียนต่อครูมีค่าเฉลี่ยเท่ากับ 21.42 และหลักสูตรการวัดและประเมินผลมีค่าเฉลี่ยเท่ากับ 0.81 โดยที่ร้อยละงบประมาณของรัฐที่จัดสรรให้มากกว่าค่าเฉลี่ยและสูงกว่าโค้งปกติ ส่วนหลักสูตร

การวัดและประเมินผลมากกว่าค่าเฉลี่ยและต่ำกว่าโค้งปกติ ส่วนทรัพยากรการเรียนในโรงเรียน และสัดส่วนนักเรียนต่อครู การกระจายข้อมูลน้อยกว่าค่าเฉลี่ยและสูงกว่าโค้งปกติ

ส่วนคะแนนความสามารถรายบุคคลของนักเรียน (θ) ในระดับนักเรียนก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า คะแนนความสามารถทั้งเพศหญิงและเพศชายคะแนนความสามารถต่ำกว่าค่าเฉลี่ยและมีการกระจายของคะแนนต่ำกว่าโค้งปกติ **ส่วนการใช้เวลาเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา** พบว่า คะแนนความสามารถของนักเรียนในกลุ่มใช้เวลาเรียนพิเศษมากกว่า 6 ชม. กลุ่มใช้เวลาเรียนพิเศษน้อยกว่า 2 ชม. และกลุ่มไม่มีเวลาเรียนเรียนพิเศษ มีคะแนนความสามารถต่ำกว่าค่าเฉลี่ยและมีการกระจายของคะแนนต่ำกว่าโค้งปกติ ส่วนกลุ่มใช้เวลาเรียนพิเศษ 2-4 ชม. และกลุ่มใช้เวลาเรียนพิเศษ 4-6 ชม. มีคะแนนความสามารถสูงกว่าค่าเฉลี่ยและมีการกระจายของคะแนนต่ำกว่าโค้งปกติ **ระดับสถานศึกษา** พบว่า สถานศึกษาขนาดใหญ่มีคะแนนความสามารถต่ำกว่าค่าเฉลี่ยและมีการกระจายของคะแนนเป็นโค้งปกติ ส่วนสถานศึกษาขนาดเล็กและขนาดกลางมีคะแนนความสามารถต่ำกว่าค่าเฉลี่ยและมีการกระจายของคะแนนต่ำกว่าโค้งปกติ **ประเภทของศึกษา** พบว่า คะแนนความสามารถของนักเรียนจากโรงเรียนเอกชนสูงกว่าค่ำเฉลี่ยและมีการกระจายของคะแนนต่ำกว่าโค้งปกติ ส่วนคะแนนความสามารถของนักเรียนจากโรงเรียนรัฐบาลต่ำกว่าค่าเฉลี่ยและมีการกระจายของคะแนนต่ำกว่าโค้งปกติ **สังกัดของสถานศึกษา** พบว่า นักเรียนที่เรียนอยู่สังกัดโรงเรียนสาธิตของมหาวิทยาลัย (สาธิต) มีคะแนนสูงกว่าค่าเฉลี่ยและ มีการกระจายของคะแนนใกล้เคียงโค้งปกติ ส่วนโรงเรียนในสังกัดกรมสามัญศึกษาเดิม (สพฐ.2) สังกัดสำนักบริหารงานคณะกรรมการส่งเสริมการศึกษาเอกชน(สช.) สังกัดสำนักการศึกษากรุงเทพมหานคร (กทม.) สังกัดสำนักประสานและพัฒนาการจัดการศึกษาท้องถิ่น (กศท.) โรงเรียนอาชีวศึกษาของเอกชน (อศ.1) วิทยาลัยอาชีวศึกษาของรัฐ (อศ.2) มีคะแนนต่ำว่าค่าเฉลี่ยและมีการกระจายของคะแนนต่ำกว่าโค้งปกติ ส่วนนักเรียนที่เรียนในโรงเรียนสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานที่มาจากโรงเรียนขยายโอกาสเดิม (สพฐ.1) มีคะแนนต่ำกว่าค่าเฉลี่ยและมีการกระจายของคะแนนสูงกว่าโค้งปกติ **ตำแหน่งที่ตั้งของสถานศึกษา** พบว่า คะแนนความสามารถของนักเรียนที่มีตำแหน่งที่ตั้งของสถานศึกษาอยู่ในหมู่บ้าน เมืองขนาดเล็ก และเมืองใหญ่มีคะแนนความสามารถต่ำกว่าค่าเฉลี่ยและมีการกระจายของคะแนนต่ำกว่าโค้งปกติ ส่วนคะแนนความสามารถของนักเรียนที่มีตำแหน่งที่ตั้งของสถานศึกษาอยู่ในเมืองและเมืองใหญ่มากมีคะแนนความสามารถต่ำกว่าค่าเฉลี่ยและมีการกระจายของคะแนนสูงกว่าโค้งปกติ **ระดับการขาดแคลนครูวิทยาศาสตร์** พบว่า คะแนนความสามารถของนักเรียนในโรงเรียนที่ไม่มีการขาดแคลนครูวิทยาศาสตร์ ระดับขาดแคลนน้อย ระดับขาดแคลนเป็นบางส่วน และระดับขาดแคลนจำนวนมากมีคะแนนต่ำกว่าค่าเฉลี่ยและมีการกระจายของคะแนนต่ำกว่าโค้งปกติ

ประเด็นที่ 2 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและของแบบสอบเมื่อใช้โมเดลมูลค่าเพิ่มพหุระดับที่ต่างกัน

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) และแบบสอบ (DTF) การทำหน้าที่ต่างกันเป็นขั้น (DSF) ซึ่งเป็นการวิเคราะห์หาอิทธิพลของตัวแปรระดับนักเรียนที่มีผลต่อโอกาสการตอบข้อสอบถูกในข้อสอบที่มีการตรวจให้คะแนนทวิวิภาคและพหุวิภาค จาก 5 ตัวแปร ได้แก่ เพศของนักเรียน (GENDER) การใช้เวลาเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา (EXTRA) เศรษฐานะของครอบครัว (ESCS) แหล่งทรัพยากรการศึกษาที่บ้าน (HEDRES) และความมั่งคั่งของครอบครัว (WEALTH) โดยยึดตามโมเดลก่อนและหลังพิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ ทั้งนี้ โมเดลหลังพิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ ใช้เกณฑ์การตัดข้อสอบที่ทำหน้าที่ต่างกันตามที่กำหนดไว้ คือ 1) ข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันมาจากตัวแปรคุณนักเรียนมากกว่าหนึ่งตัวแปร 2) เมื่อตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันบางข้อออกไม่ควรเกินร้อยละ 20 ของข้อสอบที่มีอยู่ในแต่ละฉบับ 3) ข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันซึ่งถูกตัดไปเมื่อพิจารณาการทำหน้าที่ต่างกันของแบบสอบควรทำให้อิทธิพลการทำหน้าที่ต่างกันของแบบสอบลดลงหรือเท่าเดิมได้ซึ่งขึ้นอยู่กับความตรงของโครงสร้างข้อสอบหลังการตัดข้อสอบออกไปแล้ว รายละเอียดผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) และแบบสอบ (DTF) มีดังต่อไปนี้

2.1 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ในข้อสอบที่มีการให้คะแนนทวิวิภาค และพหุวิภาค ซึ่งข้อสอบที่มีการให้คะแนนพหุวิภาคเป็นการศึกษาร่วมกันระหว่างการทำหน้าที่ต่างกันของข้อสอบ (DIF) กับการศึกษาการทำหน้าที่ต่างกันเป็นขั้น (DSF) โดยแยกตามสมรรถนะทางวิทยาศาสตร์ 3 ด้าน ได้แก่ การระบุประเด็นทางวิทยาศาสตร์ (ISI) การอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ (EPS) และการใช้ประจักษ์พยานทางวิทยาศาสตร์ (USE) ในแบบสอบที่มีการใช้จริง 9 ฉบับ พบว่า แบบสอบฉบับที่ 7 และ 13 พบว่า มีข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันมากที่สุด ร้อยละ 50 รองลงมาคือ แบบสอบฉบับที่ 10 ร้อยละ 48.57 แบบสอบฉบับที่ 9 และ 12 ร้อยละ 44.44 และแบบสอบฉบับที่ 4 พบว่า มีข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันน้อยที่สุด ร้อยละ 16.67

นอกจากนี้ยังพบว่าข้อสอบเอนเอียงเข้าข้างนักเรียนหญิงมากกว่านักเรียนชาย 3 ข้อ (ร้อยละ 10.34) เอนเอียงเข้าข้างนักเรียนที่เรียนพิเศษวิทยาศาสตร์นอกสถานศึกษามากกว่านักเรียนที่ไม่เรียนเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา 1 ข้อ (ร้อยละ 14.28) เอนเอียงเข้าข้างนักเรียน

นักเรียนที่มีเศรษฐานะของครอบครัวต่ำมากกว่านักเรียนนักเรียนที่มีเศรษฐานะของครอบครัวสูง 2 ข้อ (ร้อยละ 9.1) เอนเอียงเข้าข้างนักเรียนที่มีแหล่งทรัพยากรการศึกษาที่บ้านสูงมากกว่านักเรียนที่มีแหล่งทรัพยากรการศึกษาที่บ้านสูงต่ำ 5 ข้อ (ร้อยละ 29.42) และเอนเอียงเข้าข้างนักเรียนที่มีความมั่งคั่งทางครอบครัวสูงมากกว่านักเรียนที่มีความมั่งคั่งทางครอบครัวต่ำ 2 ข้อ (ร้อยละ 16.66)

เมื่อแยกตามตัวแปรตามสมรรถนะทางวิทยาศาสตร์ พบว่า สมรรถนะการระบุประเด็นทางวิทยาศาสตร์เอนเอียงเข้าข้างนักเรียนหญิงมากกว่านักเรียนชาย (ร้อยละ 13.79) เอนเอียงเข้าข้างนักเรียนที่ไม่เรียนพิเศษมากกว่านักเรียนที่เรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา (ร้อยละ 14.28) เอนเอียงเข้าข้างนักเรียนที่มีเศรษฐานะต่ำมากกว่านักเรียนที่มีเศรษฐานะสูง (ร้อยละ 9.09) เอนเอียงไปทางนักเรียนที่มีทรัพยากรครอบครัวต่ำมากกว่านักเรียนที่มีทรัพยากรครอบครัวสูง (ร้อยละ 11.77) และเอนเอียงเข้าข้างนักเรียนที่มีความมั่งคั่งครอบครัวต่ำมากกว่านักเรียนที่มีความมั่งคั่งครอบครัวสูง (ร้อยละ 8.33)

การอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์เอนเอียงเข้าข้างนักเรียนชายมากกว่านักเรียนหญิง (ร้อยละ 13.79) เอนเอียงเข้าข้างนักเรียนที่มีเศรษฐานะต่ำมากกว่านักเรียนที่มีเศรษฐานะสูง (ร้อยละ 4.55) เอนเอียงเข้าข้างนักเรียนที่มีทรัพยากรครอบครัวสูงมากกว่านักเรียนที่มีทรัพยากรครอบครัวต่ำ (ร้อยละ 11.77) และเอนเอียงเข้าข้างนักเรียนที่มีความมั่งคั่งครอบครัวต่ำมากกว่านักเรียนที่มีความมั่งคั่งครอบครัวสูง (ร้อยละ 8.34) แต่ไม่เอนเอียงเข้าข้างตัวแปรนักเรียนที่เรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา

ส่วนการใช้ประจักษ์พยานทางวิทยาศาสตร์เอนเอียงเข้าข้างนักเรียนที่เรียนพิเศษมากกว่านักเรียนที่ไม่เรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา (ร้อยละ 28.57) เอนเอียงเข้าข้างนักเรียนที่มีเศรษฐานะสูงมากกว่านักเรียนที่มีเศรษฐานะต่ำ (ร้อยละ 4.55) เอนเอียงเข้าข้างนักเรียนที่มีทรัพยากรครอบครัวสูงมากกว่านักเรียนที่มีทรัพยากรครอบครัวต่ำ (ร้อยละ 29.41) และเอนเอียงเข้าข้างนักเรียนที่มีความมั่งคั่งครอบครัวสูงมากกว่านักเรียนที่มีความมั่งคั่งครอบครัวต่ำ (ร้อยละ 33.33) แต่ไม่เอนเอียงเข้าข้างในตัวแปรเพศของนักเรียน

2.2 ผลการตรวจสอบการทำหน้าที่ต่างกันของแบบสอบ

ผลการตรวจสอบการทำหน้าที่ต่างกันของแบบสอบ (DTF) ในแบบสอบทั้ง 9 ฉบับที่มีการทดสอบจริงตามโครงการประเมินผลนักเรียนนานาชาติ (PISA) ปี ค.ศ. 2009 สำหรับแบบสอบที่มีการให้คะแนนแบบผสมผสาน โดยการตรวจสอบความแปรปรวนอิทธิพลการทำหน้าที่ต่างกันของข้อสอบ (the DIF effect variance) แบ่งเป็น 3 ระดับด้วยกัน คือ อิทธิพลขนาดเล็ก (S = small DIF effect variance) อิทธิพลขนาดกลาง (M = medium DIF effect variance) และ

อิทธิพลขนาดใหญ่ (L = large DIF effect variance) เกณฑ์ในการศึกษาการเกิดแบบสอบที่ตรวจพบการทำหน้าที่ต่างกันโดยพิจารณาจากอิทธิพลขนาดกลางและขนาดใหญ่ พบว่า แบบสอบที่ตรวจพบการทำหน้าที่ต่างกันมากที่สุดจำนวน 1 ฉบับ คือ แบบสอบฉบับที่ 3 (ร้อยละ 60) รองลงมาของแบบสอบที่ตรวจพบการทำหน้าที่ต่างกันจำนวน 2 ฉบับ คือ แบบสอบฉบับที่ 12 และ 13 (ร้อยละ 40 เท่ากัน) และมีจำนวน 2 ฉบับ คือแบบสอบฉบับที่ 7 และ 10 (ร้อยละ 20 เท่ากัน) ส่วนแบบสอบที่ตรวจพบการทำหน้าที่ต่างกันในระดับอิทธิพลเล็กที่สุดมีจำนวน 4 ฉบับ คือ แบบสอบฉบับที่ 2, 4, 8, และ 9 (ร้อยละ 0 เท่ากัน)

เมื่อพิจารณาตามตัวแปรที่ศึกษาโดยพิจารณาจากแบบสอบที่ตรวจพบการทำหน้าที่ต่างกันอิทธิพลขนาดกลางและขนาดใหญ่ พบว่า เศรษฐานะของครอบครัวมีแบบสอบที่ตรวจพบการทำหน้าที่ต่างกันมากที่สุดจำนวน 4 ฉบับ คือ แบบสอบฉบับที่ 3, 7, 10, และ 12 รองลงมาคือความมั่งคั่งของครอบครัวมีจำนวน 3 ฉบับ คือ แบบสอบฉบับที่ 3, 12, และ 13 และแหล่งทรัพยากรการศึกษาที่บ้านมีจำนวน 2 ฉบับ คือ แบบสอบฉบับที่ 3 และ 13 ส่วนแบบสอบที่ตรวจพบการทำหน้าที่ต่างกันในระดับอิทธิพลเล็กที่สุด คือ เพศและการใช้เวลาเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา

2.3 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและแบบสอบภายหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน

ข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันถูกตัดตามตัวแปรของนักเรียนโดยจำแนกเป็นข้อสอบที่มีการตรวจให้คะแนนทวิวิภาคและพหุวิภาค จากแบบสอบจำนวน 6 ฉบับ คือ แบบสอบฉบับที่ 7, 8, 9, 10, 1,2 และ 13 ได้ตัดข้อสอบออกไปทั้งข้อสอบที่มีการตรวจให้คะแนนทวิวิภาคและพหุวิภาค ส่วนแบบสอบจำนวน 3 ฉบับ คือ แบบสอบฉบับที่ 2, 3, และ 4 ได้ตัดข้อสอบออกไปเฉพาะข้อสอบที่มีการตรวจให้คะแนนพหุวิภาค ทั้งนี้แบบสอบฉบับที่ 2 และ 3 ตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันจำนวน 3 ตัวแปร ได้แก่ เศรษฐานะของครอบครัว แหล่งทรัพยากรการศึกษาที่บ้าน และความมั่งคั่งของครอบครัว ส่วนแบบสอบฉบับที่ 7 และ 8 ตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันจำนวน 4 ตัวแปร ได้แก่ เพศ เศรษฐานะของครอบครัว แหล่งทรัพยากรการศึกษาที่บ้าน และความมั่งคั่งของครอบครัว ส่วนแบบสอบฉบับที่ 13 ตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันจำนวน 4 ตัวแปร ได้แก่ การเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา เศรษฐานะของครอบครัว แหล่งทรัพยากรการศึกษาที่บ้าน และความมั่งคั่งของครอบครัว และแบบสอบฉบับที่ 4, 9, 10, และ 12 ตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันมี

ครบทั้ง 5 ตัวแปร ได้แก่ เพศ การเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา เศรษฐานะของครอบครัว แหล่งทรัพยากรการศึกษาที่บ้าน และความมั่งคั่งของครอบครัว

นอกจากนี้เมื่อตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันออกไปในแบบสอบแต่ละฉบับโดยยึดเกณฑ์การตัดข้อสอบไม่เกินร้อยละ 20 พบว่า แบบสอบจำนวน 5 ฉบับ คือ ฉบับที่ 4, 7, 9, 12, และ 13 มีร้อยละการตัดจำนวนข้อสอบมากที่สุด (ร้อยละ 16.67 เท่ากัน) รองลงมาคือแบบสอบจำนวน 2 ฉบับ คือ แบบสอบฉบับที่ 3 และ 10 (ร้อยละ 14.29 เท่ากัน) และน้อยที่สุดคือแบบสอบจำนวน 2 ฉบับ คือ แบบสอบฉบับที่ 2 และ 8 (ร้อยละ 11.76 เท่ากัน) และเมื่อตรวจสอบค่าความเที่ยงของข้อสอบก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า แบบสอบทุกฉบับมีค่าความเที่ยงตั้งแต่ .686 - .857 โดยมีแบบสอบฉบับที่ 12 มีค่าความเที่ยงมากที่สุด (.857) รองลงมาคือแบบสอบฉบับที่ 3 และ 10 ตามลำดับ (.844, .809 ตามลำดับ) และแบบสอบฉบับที่ 4 มีค่าความเที่ยงน้อยที่สุด (.686) และค่าความเที่ยงของข้อสอบหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า แบบสอบทุกฉบับมีค่าความเที่ยงตั้งแต่ .620 - .821 โดยแบบสอบฉบับที่ 12 มีค่าความเที่ยงมากที่สุด (.821) รองลงมาคือแบบสอบฉบับที่ 3 และ 10 ตามลำดับ (.809, .785 ตามลำดับ) และแบบสอบฉบับที่ 4 มีค่าความเที่ยงน้อยที่สุด (.620)

ผลการเปรียบเทียบร้อยละแบบสอบที่ตรวจพบการทำหน้าที่ต่างกันทั้งก่อนและหลังตัดที่ตรวจพบการทำหน้าที่ต่างกัน โดยใช้เกณฑ์การพิจาณาจากอิทธิพลแบบสอบที่ตรวจพบการทำหน้าที่ต่างกันขนาดกลาง (M) และขนาดใหญ่ (L) พบว่า ร้อยละแบบสอบที่ตรวจพบการทำหน้าที่ต่างกันลดลงในแบบสอบฉบับที่ 12 ลดลงจากร้อยละ 40 เป็น 20 และแบบสอบฉบับที่ 13 จากร้อยละ 20 เป็น 0 ส่วนแบบสอบฉบับที่ 3, 7, และ 10 ยังคงมีร้อยแบบสอบที่ตรวจพบการทำหน้าที่ต่างกันเท่าเดิม และเมื่อพิจารณาตามตัวแปรที่ศึกษา พบว่า แบบสอบฉบับที่ 13 มี 2 ตัวแปร คือ แหล่งทรัพยากรการศึกษาที่บ้านและความมั่งคั่งของครอบครัวมีอิทธิพลแบบสอบทำหน้าที่ต่างกันลดลงจากขนาดกลางเป็นขนาดเล็ก และแบบสอบฉบับที่ 12 มี 1 ตัวแปร คือความมั่งคั่งของครอบครัวมีอิทธิพลแบบสอบทำหน้าที่ต่างกันลดลงจากขนาดกลางเป็นขนาดเล็ก

ต่อมาผู้วิจัยได้ตรวจสอบความตรงเชิงโครงสร้างด้วยการวิเคราะห์องค์ประกอบเชิงยืนยันเพื่อทดสอบความสอดคล้องของโมเดลตามสมรรถนะทางวิทยาศาสตร์ โดยการยึดองค์ประกอบตามสมรรถนะทางวิทยาศาสตร์ 3 ด้าน ก่อนที่จะทำการวิเคราะห์องค์ประกอบเชิงยืนยัน ผู้วิจัยได้ศึกษาความสัมพันธ์ระหว่างองค์ประกอบทั้ง 3 ด้านในแบบสอบแต่ละฉบับและภาพรวมทุกฉบับ เพื่อพิจารณาความเหมาะสมของเมทริกสหสัมพันธ์ที่จะนำไปสู่การวิเคราะห์องค์ประกอบ โดยใช้สถิติของ Bartlett เป็นสถิติที่ใช้ทดสอบสมมติฐานความเป็นเมทริกเอกลักษณ์ (identity matrix) และค่าดัชนีไกเซอร์-ไมเยอร์-ออลคิน (Kaiser-Meyer-Olkin Measure of Sampling Adequacy=MSA) เพื่อทดสอบความเหมาะสมของข้อมูล พบว่า แบบสอบภาพรวมทุกฉบับและรายฉบับก่อนการตัด

ข้อสอบที่ตรวจพบการทำหน้าหน้าที่ต่างกันมีเมทริกสหสัมพันธ์ระหว่างตัวบ่งชี้แตกต่างจากเมทริกเอกลักษณ์อย่างมีนัยสำคัญทางสถิติ และค่าดัชนีไกเซอร์-ไมเยอร์-ออลคิน มีค่าตั้งแต่ .566 - .709 ซึ่งมากกว่า .5 มีความเหมาะสมที่จะนำมาวิเคราะห์องค์ประกอบ ส่วนแบบสอบภาพรวมทุกฉบับและรายฉบับหลังตัดข้อสอบตรวจพบการทำหน้าหน้าที่ต่างกัน พบว่า เมทริกสหสัมพันธ์ระหว่างตัวบ่งชี้แตกต่างจากเมทริกเอกลักษณ์อย่างมีนัยสำคัญทางสถิติ และค่าดัชนีไกเซอร์-ไมเยอร์-ออลคิน มีค่าตั้งแต่ .524 - .708 มากกว่า .5 หมายความว่า แบบสอบภาพรวมทุกฉบับและรายฉบับภายหลังตัดข้อสอบที่ตรวจพบการทำหน้าหน้าที่ต่างกันไปแล้วยังคงมีความตรงเชิงโครงสร้างตามองค์ประกอบสมรรถนะทางวิทยาศาสตร์

ประเด็นที่ 3 ผลการเปรียบเทียบประสิทธิภาพโมเดลมูลค่าเพิ่มการจัดการศึกษาวิทยาศาสตร์ระหว่าง 2 โมเดล

จากผลการเปรียบเทียบประสิทธิภาพโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ ผู้วิจัยได้สรุปผลการเปรียบเทียบประสิทธิภาพโมเดลมูลค่าเพิ่มการจัดการศึกษาออกเป็น 5 ส่วน ได้แก่ ส่วนที่หนึ่ง คือ ผลการตรวจสอบเงื่อนไขการวิเคราะห์พหุระดับเบื้องต้น ส่วนที่สอง คือ ผลการเปรียบเทียบประสิทธิภาพโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ ส่วนที่สาม คือ ผลการศึกษาความสัมพันธ์ระหว่างโมเดลมูลค่าเพิ่มพหุระดับของ 2 โมเดล ส่วนที่สี่ คือ ผลการเปรียบเทียบความสอดคล้องของประสิทธิผลของโรงเรียน การจัดอันดับคุณภาพภาพและการจัดกลุ่มคุณภาพระหว่างโมเดลมูลค่าเพิ่มพหุระดับระหว่าง 2 โมเดล รายละเอียดดังต่อไปนี้

3.1 ผลการตรวจสอบเงื่อนไขการวิเคราะห์พหุระดับเบื้องต้น

ผลการตรวจสอบเงื่อนไขการวิเคราะห์พหุระดับเบื้องต้นโดยการตรวจสอบความสัมพันธ์ระหว่างตัวแปรจากการตรวจสอบภาวะร่วมเส้นตรงพหุ (multicollinearity) และการตรวจสอบความสอดคล้องกับสมการการวิเคราะห์ถดถอยและโมเดลพหุระดับ พบว่า ผลการวิเคราะห์ค่าสหสัมพันธ์ของตัวแปรในระดับนักเรียนก่อนตัดข้อสอบที่ตรวจพบทำหน้าที่ต่างกัน ได้แก่ คะแนนสเกลประเมินผลการรู้เรื่องวิทยาศาสตร์ก่อนตัดข้อสอบที่ตรวจพบทำหน้าที่ต่างกัน ตัวแปรดัมมี่ความเป็นเพศหญิงของนักเรียน ตัวแปรดัมมี่การใช้เวลาเรียนในการเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา เศรษฐานะของครอบครัว แหล่งทรัพยากรการศึกษาที่บ้าน ความ มั่งคั่งของครอบครัว และแหล่งทรัพยากรทางไอซีที เมื่อพิจารณาความสัมพันธ์ของตัวแปรระดับนักเรียน พบว่า ตัวแปรตาม คือ คะแนนสเกลประเมินผลการรู้เรื่องวิทยาศาสตร์ก่อนตัดข้อสอบที่

ตรวจพบทำหน้าที่ต่างกัน มีความสัมพันธ์กับตัวแปรทำนายทั้ง 6 ตัวแปรอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

ส่วนผลการวิเคราะห์ค่าสหสัมพันธ์ของตัวแปรในระดับนักเรียนหลังตัดข้อสอบที่ตรวจพบทำหน้าที่ต่างกัน ได้แก่ คะแนนสเกลประเมินผลการรู้เรื่องวิทยาศาสตร์หลังตัดข้อสอบที่ตรวจพบทำหน้าที่ต่างกัน ตัวแปรดัมมี่ความเป็นเพศหญิงของนักเรียน ตัวแปรดัมมี่การใช้เวลาเรียนในการเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา เศรษฐานะของครอบครัว แหล่งทรัพยากรการศึกษาที่บ้าน ความมั่งคั่งของครอบครัว และแหล่งทรัพยากรทางไอซีที เมื่อพิจารณาความสัมพันธ์ของตัวแปรระดับนักเรียน พบว่า ตัวแปรตามคือ คะแนนสเกลประเมินผลการรู้เรื่องวิทยาศาสตร์หลังตัดข้อสอบที่ตรวจพบทำหน้าที่ต่างกัน มีความสัมพันธ์กับตัวแปรทำนายทั้ง 6 ตัวแปร อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 เช่นเดียวกัน

ผลการตรวจสอบภาวะร่วมเส้นตรงพหุ (multicollinearity) ก่อนและหลังตัดข้อสอบที่ตรวจพบทำหน้าที่ต่างกันในตัวแปรระดับนักเรียน โดยการวิเคราะห์ถดถอยพหุ (multiple regression) มีตัวแปรตามเป็นคะแนนสเกลประเมินผลการรู้เรื่องวิทยาศาสตร์ก่อนและหลังตัดข้อสอบที่ตรวจพบทำหน้าที่ต่างกัน และตัวแปรทำนาย ได้แก่ ตัวแปรดัมมี่ความเป็นเพศหญิงของนักเรียน ตัวแปรดัมมี่การใช้เวลาเรียนในการเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา เศรษฐานะของครอบครัว แหล่งทรัพยากรการศึกษาที่บ้าน ความมั่งคั่งของครอบครัว และแหล่งทรัพยากรทางไอซีที โดยพิจารณาค่าความคลาดเคลื่อนยินยอม (Tolerance) และค่า VIF (variance inflation factor) ซึ่งค่าความคลาดเคลื่อนยินยอม (Tolerance) ควรสูงกว่า 0.1 และค่า VIF ควรมีค่าน้อยกว่า 10 (Hair et al., 1998) ส่วนค่าดัชนีเงื่อนไข (condition index) ไม่ควรเกิน 30 (สมถวิล วิจิตรวรรณา, สุภมาส อังศุโชติ และรัชนีกูล ภิญโญภานุวัฒน์, 2553) พบว่า ไม่เกิดปัญหาตัวแปรทำนายมีความสัมพันธ์กันสูงหรือไม่เกิดภาวะร่วมเส้นตรงพหุก่อนและหลังการตัดข้อสอบทำหน้าที่ต่างกันของตัวแปรทำนายทั้ง 6 ตัวแปร ซึ่งผลที่ได้มีค่าความคลาดเคลื่อนยินยอม (Tolerance) สูงกว่า 0.1 และค่า VIF น้อยกว่า 10 ส่วนค่าดัชนีเงื่อนไข (condition index) ทั้ง 6 ตัวแปร ไม่เกิน 30 ดังนั้นจึงไม่ต้องตัดตัวแปรทำนายระดับนักเรียนตัวใดทิ้ง

ส่วนผลการตรวจสอบภาวะร่วมเส้นตรงพหุ (multicollinearity) ก่อนและหลังตัดข้อสอบที่ตรวจพบทำหน้าที่ต่างกันในตัวแปรระดับสถานศึกษา มีตัวแปรตามเป็นคะแนนเฉลี่ยสเกลประเมินผลการรู้เรื่องวิทยาศาสตร์ก่อนและหลังตัดข้อสอบที่ตรวจพบทำหน้าที่ต่างกัน และตัวแปรทำนาย ได้แก่ ตัวแปรดัมมี่สถานศึกษาของรัฐ ระดับการขาดแคลนครู ร้อยละงบประมาณของรัฐที่จัดสรรให้สถานศึกษา ดัชนีทรัพยากรการเรียน สัดส่วนนักเรียนต่อครู และหลักสูตรการวัดและประเมินผล พบว่า ไม่เกิดปัญหาตัวแปรทำนายมีความสัมพันธ์กันสูงหรือไม่เกิดภาวะร่วมเส้นตรงพหุก่อนและหลังการตัดข้อสอบทำหน้าที่ต่างกันของตัวแปรทำนายทั้ง 6 ตัวแปร ซึ่งผลที่ได้มีค่าความคลาดเคลื่อนยินยอม

(Tolerance) สูงกว่า 0.1 และค่า VIF น้อยกว่า 10 ส่วนค่าดัชนีเงื่อนไข (condition index) ทั้ง 6 ตัวแปร ไม่เกิน 30 ดังนั้นจึงไม่ต้องตัดตัวแปรทำนายระดับสถานศึกษาตัวใดทิ้ง

3.2 ผลการเปรียบเทียบประสิทธิภาพโมเดลการประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์

ผลการวิเคราะห์โมเดลมูลค่าเพิ่มพหุระดับซึ่งเป็นโมเดลตามวัตถุประสงค์การวิจัยในโมเดล Undetected DIF-DSF-DTF ที่มีการควบคุมตัวแปรระดับนักเรียนและสถานศึกษาผลการทดสอบอิทธิพลคงที่ (fixed effect) พบว่า ค่าเฉลี่ยรวมคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์ทุกสถานศึกษา (G_{00}) มีความผันแปรจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 (G_{00} = 465.762, t = 58.115, p = .000) ค่าสัมประสิทธิ์ถดถอยของตัวแปรระดับสถานศึกษาที่มีค่าเป็นบวกสูงที่สุด คือ สถานศึกษาของรัฐบาล (β =36.595) รองลงมาคือ ดัชนีทรัพยากรการเรียน (β = 12.413) ส่วนค่าสัมประสิทธิ์ถดถอยของตัวแปรระดับสถานศึกษาที่มีค่าเป็นลบสูงที่สุด คือ ร้อยละงบประมาณของรัฐที่จัดสรรให้สถานศึกษา (β = -0.414) ส่วนตัวแปรระดับนักเรียนที่มีค่าเป็นบวกสูงที่สุด คือ เศรษฐานะของครอบครัว (β = 13.183) รองลงมาคือ แหล่งทรัพยากรการศึกษาที่บ้าน (β = 10.709) ส่วนค่าสัมประสิทธิ์ถดถอยของตัวแปรระดับสถานศึกษาที่มีค่าเป็นลบสูงที่สุด คือ ความมั่งคั่งของครอบครัว (β = -16.088)

ผลการทดสอบอิทธิพลเชิงสุ่ม (random effect) พบว่า ค่าเศษเหลือระดับสถานศึกษาของคะแนนการประเมินผลการรู้เรื่องวิทยาศาสตร์ที่มีการควบคุมตัวแปรระดับนักเรียนและสถานศึกษา (U_{0k}) หรือมูลค่าเพิ่มของสถานศึกษาที่ความผันแปรระหว่างสถานศึกษาอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 (χ^2 = 684.761, p = .000) เมื่อพิจารณาเมทริกของความแปรปรวนร่วมหรือสหสัมพันธ์ภายในชั้น (intra-class correlation) ของโมเดลที่ 3 เท่ากับ 0.161 หมายความว่า 16.1% เป็นความแปรปรวนของคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์ระหว่างโรงเรียน อีก 83.9% เป็นความแปรปรวนของคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์ระหว่างนักเรียน การประมาณค่าความเที่ยงของสัมประสิทธิ์ตัวแปรสุ่มในระดับที่ 1 มีค่าเท่ากับ 0.614

ส่วนโมเดลตามวัตถุประสงค์การวิจัยในโมเดล Detected DIF-DSF-DTF ที่มีการควบคุมตัวแปรระดับนักเรียนและสถานศึกษา ผลการทดสอบอิทธิพลคงที่ (Fixed effect) พบว่า ค่าเฉลี่ยรวมคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์ทุกสถานศึกษา (G_{00}) มีความผันแปรจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 (G_{00} = 469.839, t = 60.135, p = .000) ค่าสัมประสิทธิ์ถดถอยของตัวแปรระดับสถานศึกษาที่มีค่าเป็นบวกสูงที่สุดคือ สถานศึกษาที่เป็นสถานศึกษาของรัฐบาล (β = 33.688074) รองลงมาคือ ทรัพยากรการเรียน (β = 11.247) ส่วนค่าสัมประสิทธิ์ถดถอยของ

ตัวแปรระดับสถานศึกษาที่มีค่าเป็นลบสูงที่สุด คือ ร้อยละงบประมาณของรัฐที่จัดสรรให้สถานศึกษา (β = -0.379) ส่วนตัวแปรระดับนักเรียนที่มีค่าเป็นบวกสูงที่สุดคือ เศรษฐานะของครอบครัว (β =13.025) รองลงมาคือ แหล่งทรัพยากรการศึกษาที่บ้าน (β = 9.745) ส่วนค่าสัมประสิทธิ์ถดถอยของตัวแปรระดับสถานศึกษาที่มีค่าเป็นลบสูงที่สุด คือ ความมั่งคั่งของครอบครัว (β =-15.789)

ผลการทดสอบอิทธิพลเชิงสุ่ม (random effect) พบว่า ค่าเศษเหลือระดับสถานศึกษาของคะแนนการประเมินผลการรู้เรื่องวิทยาศาสตร์ที่มีการควบคุมตัวแปรระดับนักเรียนและสถานศึกษา (U_{0k}) หรือมูลค่าเพิ่มของสถานศึกษาที่ความผันแปรระหว่างสถานศึกษาอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 (χ^2=653.075, p=.000) เมื่อพิจารณาเมทริกของความแปรปรวนร่วมหรือสหสัมพันธ์ภายในชั้น (intra-class correlation) ของโมเดลที่ 4 เท่ากับ 0.151 หมายความว่า 15.1% เป็นความแปรปรวนของคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์ระหว่างโรงเรียน อีก 84.9% เป็นความแปรปรวนของคะแนนผลการประเมินการรู้เรื่องวิทยาศาสตร์ระหว่างนักเรียน การประมาณค่าความเที่ยงของสัมประสิทธิ์ตัวแปรสุ่มในระดับที่ 1 มีค่าเท่ากับ 0.599

เมื่อคำนวณสัมประสิทธิ์ทำนายหรืออธิบายได้ (R^2) ของโมเดล Undetected DIF-DSF-DTF พบว่า กลุ่มตัวแปรระดับนักเรียนและกลุ่มตัวแปรบริบทระดับสถานศึกษาสามารถอธิบายสัดส่วนความแปรปรวนได้ถึง .49507 (49.507%) ส่วนโมเดล Detected DIF-DSF-DTF พบว่า กลุ่มตัวแปรระดับนักเรียนและกลุ่มตัวแปรบริบทระดับสถานศึกษาสามารถอธิบายสัดส่วนความแปรปรวนได้ถึง .49946 (49.946%) หมายความว่า ค่าเฉลี่ยรวมคะแนนการประเมินผลการรู้เรื่องวิทยาศาสตร์ที่มีการควบคุมตัวแปรระดับนักเรียนและสถานศึกษาหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันสามารถอธิบายสัดส่วนความแปรปรวนได้มากกว่าข้อสอบก่อนข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน รายละเอียดตามตาราที่ 61

Chulalongkorn University

ตารางที่ 61 สรุปสหสัมพันธ์ภายในชั้นและสัมประสิทธิ์การทำนาย

ตัวแปร	ประเภท Model	
	Undetected DIF-DSF-DTF	Detected DIF-DSF-DTF
สหสัมพันธ์ภายในชั้น (Intra-class correlation)		
ระหว่างโรงเรียน	16.1%	15.1%
ระหว่างนักเรียน	83.9%	84.9%
สัมประสิทธิ์การทำนายหรืออธิบายได้ (R^2) $= \frac{(\hat{\tau}_{00}(\text{uncon}) - \hat{\tau}_{00}(\text{hypo})}{\hat{\tau}_{00}(\text{uncon})}$	.49507 (49.507%)	.49946 (49.946%)

3.3 ผลการศึกษาความสัมพันธ์ระหว่างโมเดลมูลค่าเพิ่มพหุระดับของ 2 โมเดล

ในส่วนผลการศึกษาความสัมพันธ์ระหว่างโมเดลมูลค่าเพิ่มพหุระดับโมเดล Undetected DIF-DSF-DTF (โมเดลที่ 1) และ โมเดล Detected DIF-DSF-DTF (โมเดลที่ 2) ผู้วิจัยแบ่งการนำเสนอออกเป็นการวิเคราะห์ความสัมพันธ์ระหว่างคะแนนเฉลี่ยการประเมินผลการรู้เรื่องวิทยาศาสตร์ คะแนนมูลค่าเพิ่ม (value-added scores: VAS) การจัดอันดับคุณภาพคะแนนมูลค่าเพิ่ม (value-added scores ranking: VAS ranking) และการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม (value-added scores rating: VAS rating) และการวิเคราะห์ความสัมพันธ์ระหว่างคะแนนความสามารถ (theta scores) และคะแนนมูลค่าเพิ่ม (value-added scores) ตามตัวแปรที่ศึกษา ดังนี้

3.3.1 ผลการวิเคราะห์ความสัมพันธ์ระหว่างคะแนนเฉลี่ยการประเมินผลการรู้เรื่องวิทยาศาสตร์ คะแนนมูลค่าเพิ่ม การจัดอันดับคุณภาพคะแนนมูลค่าเพิ่ม และการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม

ผลการวิเคราะห์ความสัมพันธ์ในโมเดลที่ 1 พบว่า ความสัมพันธ์ระหว่างตัวแปรคะแนนมูลค่าเพิ่ม (VAS) กับตัวแปรการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่ม (VAS ranking) ($r=-.966$, $p<.01$) มีความสัมพันธ์เกือบสมบูรณ์ในทิศทางตรงกันข้าม ความสัมพันธ์ระหว่างตัวแปรคะแนนมูลค่าเพิ่ม (VAS) กับตัวแปรการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม (VAS rating) ($r=.946$, $p<.01$) มีความสัมพันธ์เกือบสมบูรณ์ในทิศทางเดียวกัน ส่วนความสัมพันธ์ระหว่างตัวแปรการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่ม (VAS ranking) กับตัวแปรการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม (VAS rating) ($r=-.926$, $p<0.01$) มีความสัมพันธ์เกือบสมบูรณ์ในทิศทางตรงกันข้าม

ส่วนผลการวิเคราะห์ความสัมพันธ์ในโมเดลที่ 2 พบว่า ความสัมพันธ์ระหว่างตัวแปรคะแนนมูลค่าเพิ่ม (VAS) กับตัวแปรการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่ม (VAS ranking) ($r=-.963$, $p<.01$) มีความสัมพันธ์เกือบสมบูรณ์ในทิศทางตรงกันข้าม ความสัมพันธ์ระหว่างตัวแปรคะแนนมูลค่าเพิ่ม (VAS) กับตัวแปรการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม (VAS rating) ($r=.945$, $p<.01$) มีความสัมพันธ์เกือบสมบูรณ์ในทิศทางเดียวกัน ส่วนความสัมพันธ์ระหว่างตัวแปรการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่ม (VAS ranking) กับตัวแปรการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม (VAS rating) ($r=-.926$, $p<0.01$) มีความสัมพันธ์เกือบสมบูรณ์ในทิศทางตรงกันข้าม

3.3.2 การวิเคราะห์ความสัมพันธ์ระหว่างคะแนนความสามารถและคะแนนมูลค่าเพิ่มตามตัวแปรที่ศึกษา

ผลการวิเคราะห์ความสัมพันธ์ระหว่างคะแนนความสามารถ (theta scores) และคะแนนมูลค่าเพิ่ม (value-added scores) กับ 3 ตัวแปร ได้แก่ เศรษฐานะของครอบครัว แหล่งทรัพยากรการศึกษาที่บ้าน และความมั่งคั่งของครอบครัว ซึ่งเป็นตัวแปรต่อเนื่องระดับนักเรียนที่ถูกนำมาพิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ และการทำหน้าที่ต่างกันเป็นขั้น และนำมาควบคุมในโมเดลวิเคราะห์มูลค่าเพิ่มพหุระดับระหว่างโมเดล ที่ 1 กับโมเดลที่ 2 พบว่า ความสัมพันธ์ระหว่างคะแนนความสามารถ (theta scores) กับ 3 ตัวแปร มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และความสัมพันธ์ระหว่างคะแนนมูลค่าเพิ่ม (value-added scores) กับ 3 ตัวแปร ได้แก่ เศรษฐานะของครอบครัว แหล่งทรัพยากรการศึกษาที่บ้าน และความมั่งคั่งของครอบครัว มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 เช่นเดียวกัน และเมื่อเปรียบเทียบความสัมพันธ์ระหว่างคะแนนค่าความสามารถก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันกับ 3 ตัวแปร กับคะแนนมูลค่าเพิ่มหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันกับ 3 ตัวแปร พบว่า ค่าความสัมพันธ์ระหว่างคะแนนมูลค่าเพิ่มหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันลดลง

3.4 ผลการเปรียบเทียบความสอดคล้องของประสิทธิผลของโรงเรียน การจัดอันดับคุณภาพภาพและการจัดกลุ่มคุณภาพระหว่างโมเดลมูลค่าเพิ่มพหุระดับระหว่าง 2 โมเดล

ในส่วนผลการเปรียบเทียบความสอดคล้องของประสิทธิผลของโรงเรียนการจัดอันดับคุณภาพภาพและการจัดกลุ่มคุณภาพระหว่างโมเดลมูลค่าเพิ่มพหุระดับ ระหว่างโมเดล Undetected DIF-DSF-DTF (โมเดลที่ 1) กับโมเดล Detected DIF-DSF-DTF (โมเดลที่ 2) ผู้วิจัยแบ่งการนำเสนอออกเป็น 3 ส่วน คือ ผลการเปรียบเทียบความสอดคล้องประสิทธิผลของโรงเรียน ผลการเปรียบเทียบความสอดคล้องการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่ม (VAS ranking) และผลเปรียบเทียบความสอดคล้องการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม (VAS rating) ดังนี้

3.4.1 ผลการผลการเปรียบเทียบความสอดคล้องประสิทธิผลของโรงเรียน

ผลการวิเคราะห์ พบว่า ทั้งสองโมเดลมีผลการทดสอบประสิทธิผลของโรงเรียนที่สอดคล้องกัน โดยที่โมเดลการประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ที่ไม่ได้พิจารณาข้อสอบที่ทำหน้าที่ต่างกันข้อสอบและแบบสอบ พบว่า โรงเรียนที่ไม่มีประสิทธิผลจำนวน 128 โรงเรียน

และโรงเรียนที่มีประสิทธิผลจำนวน 95 โรงเรียน ภายหลังจากที่มีการทดสอบความเท่ากันของค่าความแปรปรวน (levene's test) พบว่า มีความแปรปรวนของคะแนนมูลค่าเพิ่มระหว่างกลุ่มโรงเรียนที่มีประสิทธิผลและโรงเรียนที่ไม่มีประสิทธิผลแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 (levene statistic=18.929, p=.00) ส่วนโมเดลการประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ที่พิจารณาข้อสอบที่ทำหน้าที่ต่างกันข้อสอบและแบบสอบ พบว่า โรงเรียนที่ไม่มีประสิทธิผลจำนวน 128 โรงเรียน และโรงเรียนที่มีประสิทธิผลจำนวน 95 โรงเรียน ภายหลังจากที่มีการทดสอบความเท่ากันของค่าความแปรปรวน (levene's test) พบว่า มีความแปรปรวนของคะแนนมูลค่าเพิ่มระหว่างกลุ่มโรงเรียนที่มีประสิทธิผลและโรงเรียนที่ไม่มีประสิทธิผลแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 (levene statistic=19.133, p=.00)

3.4.2 ผลการเปรียบเทียบความสอดคล้องการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่ม

ผลการเปรียบเทียบความสอดคล้องการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่ม (VAS ranking) การจัดการศึกษาวิชาวิทยาศาสตร์ระหว่างโมเดลที่ 1 กับโมเดลที่ 2 ผลการทดสอบความสอดคล้องการจัดอันดับคุณภาพการจัดการศึกษาโดยสถิติทดสอบ Wilcoxon Signed Ranks Test พบว่า โมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ตามการจัดอันดับคุณภาพระหว่างโมเดลที่1 กับโมเดลที่ 2 ไม่สอดคล้องกัน (Z = -.219, p= .829) โดยที่อันดับคุณภาพของ โมเดลที่ 1 มีอันดับน้อยกว่า โมเดลที่ 2 จำนวน 97 โรงเรียน ส่วนอันดับคุณภาพของ โมเดลที่ 1 มีอันดับมากกว่าโมเดลที่ 2 จำนวน 107 โรงเรียน และอันดับคุณภาพของ โมเดลที่ 1 เท่ากันกับโมเดลที่ 2 จำนวน 19 โรงเรียน

CHULALONGKORN UNIVERSITY

3.4.3 ผลเปรียบเทียบความสอดคล้องการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม

ผลการทดสอบความสอดคล้องของการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม (VAS rating) การจัดการศึกษาวิชาวิทยาศาสตร์ระหว่างโมเดลที่ 1 กับโมเดลที่ 2 โดยการจัดกลุ่มคุณภาพการจัดการศึกษา 5 ระดับคือ ระดับ A (ดีมาก) ระดับ B (ดี) ระดับ C (พอใช้) ระดับ D (ผ่าน) และระดับ E (ควรปรับปรุง) ผลการทดสอบความสอดคล้องการจัดระดับคุณภาพการจัดการศึกษาโดยอาศัยหลักเกณฑ์การวัดความสอดคล้อง (Measure of Agreement) ของ Cohen's Kappa ระหว่างโมเดลที่ 1 กับโมเดลที่ 2 มีการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่มที่สอดคล้องกันระดับเกือบสมบูรณ์ (Cohen's Kappa = .865) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

โดยที่ผลการทดสอบความสอดสอดคล้องการจัดกลุ่มคุณภาพของสถานศึกษา พบว่า โมเดลที่ 1 สอดคล้องกับโมเดลที่ 2 โดยภาพรวมจำนวน 202 โรงเรียน คิดเป็นร้อยละที่สอดคล้อง

ร้อยละ 90.58 ซึ่งที่ระดับคุณภาพ E (ควรปรับปรุง) มีจำนวนสถานศึกษาสอดคล้องกัน 74 โรงเรียน ระดับคุณภาพ D (ผ่าน) มีจำนวนสถานศึกษาสอดคล้องกัน 70 โรงเรียน ระดับคุณภาพ C (พอใช้) มีจำนวนสถานศึกษาสอดคล้องกัน 39 โรงเรียน ระดับคุณภาพ B (ดี) มีจำนวนสถานศึกษาสอดคล้องกัน 17 โรงเรียน และระดับคุณภาพ A (ดีมาก) มีจำนวนสถานศึกษาสอดคล้องกัน 2 โรงเรียน

จากการทดสอบความสอดคล้องการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม ระหว่างโมเดลที่ 1 กับโมเดลที่ 2 โดยจำแนกตามตัวแปรคุณลักษณะสถานศึกษา ได้แก่ ขนาดสถานศึกษา ประเภทสถานศึกษา สังกัดของสถานศึกษา ตำแหน่งที่ตั้งของสถานศึกษา และระดับการขาดแคลนครูวิทยาศาสตร์ เป็นดังนี้

ขนาดสถานศึกษา เมื่อแยกพิจารณาเป็น**ขนาดสถานศึกษาขนาดเล็ก** พบว่า จากจำนวนสถานศึกษา 169 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 152 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 89.94 **ขนาดสถานศึกษาขนาดกลาง** พบว่า จากจำนวนสถานศึกษา 51 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 47 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 92.16 และ**ขนาดสถานศึกษาขนาดใหญ่** พบว่า จากจำนวนสถานศึกษา 3 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 3 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 100

ประเภทสถานศึกษา เมื่อแยกพิจารณาเป็น**ประเภทสถานศึกษาของรัฐ** พบว่า จากจำนวนสถานศึกษา 197 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 178 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 90.36 และ**ประเภทสถานศึกษาเอกชน** พบว่า จากจำนวนสถานศึกษา 26 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 24 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 92.31

สังกัดของสถานศึกษา เมื่อแยกพิจารณาเป็น**โรงเรียนในสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานที่มาจากโรงเรียนขยายโอกาสเดิม (สพฐ.1)** พบว่า จากจำนวนสถานศึกษา 35 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 30 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 85.71 **โรงเรียนในสังกัดกรมสามัญศึกษาเดิม (สพฐ.2)** พบว่า จากจำนวนสถานศึกษา 91 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 82 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 90.11 **โรงเรียนในสังกัดสำนักบริหารงานคณะกรรมการส่งเสริมการศึกษาเอกชน (สช.)** พบว่า จากจำนวนสถานศึกษา 13 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 13 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 100.00 **โรงเรียนในสังกัดสำนักการศึกษากรุงเทพมหานคร (กทม.)** พบว่า จากจำนวนสถานศึกษา 20 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 19 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 95.00 **โรงเรียนในสังกัดสำนักประสานและพัฒนาการจัดการศึกษาท้องถิ่น (กศท.)** พบว่า จากจำนวนสถานศึกษา 17 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 16 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 94.12 **โรงเรียนในสังกัดสาธิตของมหาวิทยาลัย (สาธิต)** พบว่า จากจำนวนสถานศึกษา

16 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 15 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 93.75 **โรงเรียนอาชีวศึกษาของเอกชน (อศ.1)** พบว่า จากจำนวนสถานศึกษา 13 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 11 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 84.62 และ**วิทยาลัยอาชีวศึกษาของรัฐ (อศ.2)** พบว่า จากจำนวนสถานศึกษา 18 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 16 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 88.89

ตำแหน่งที่ตั้งของสถานศึกษา เมื่อแยกพิจารณาเป็น**สถานศึกษาที่ตั้งอยู่ในหมู่บ้าน** พบว่า จากจำนวนสถานศึกษา 47 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 40 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 85.11 **สถานศึกษาที่ตั้งอยู่ในเมืองเล็ก** พบว่า จากจำนวนสถานศึกษา 45 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 43 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 95.56 **สถานศึกษาที่ตั้งอยู่ในเมือง** พบว่า จากจำนวนสถานศึกษา 73 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 65 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 89.04 **สถานศึกษาที่ตั้งอยู่ในเมืองใหญ่** พบว่า จากจำนวนสถานศึกษา 38 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 34 โรงเรียน คิดเป็นร้อยละที่สอดคล้อง ร้อยละ 89.47 และ**สถานศึกษาที่ตั้งอยู่ในเมืองใหญ่มาก** พบว่า จากจำนวนสถานศึกษา 20 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 20 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 100.00

ระดับการขาดแคลนครูวิทยาศาสตร์ เมื่อแยกพิจารณาเป็น**ระดับไม่ขาดแคลนครูวิทยาศาสตร์** พบว่า จากจำนวนสถานศึกษา 86 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 77 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 89.53 **ระดับขาดแคลนครูวิทยาศาสตร์น้อย** พบว่า จากจำนวนสถานศึกษา 41 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 40 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 97.56 **ระดับขาดแคลนครูวิทยาศาสตร์เป็นบางส่วน** พบว่า จากจำนวนสถานศึกษา 81 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 72 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 88.89 และ**ระดับขาดแคลนครูวิทยาศาสตร์จำนวนมาก** พบว่า จากจำนวนสถานศึกษา 15 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 13 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 86.67

5.2 อภิปรายผลการวิจัย

การวิจัยครั้งนี้เป็นการวิเคราะห์เปรียบเทียบโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์โดยการประยุกต์ใช้โมเดลมูลค่าเพิ่มพหุระดับที่มีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ ผู้วิจัยได้อภิปรายผลการวิจัยได้เป็น 2 ประเด็นหลัก ตามเนื้อหาความสำคัญที่สอดคล้องกับวัตถุประสงค์และสมมติฐานการวิจัย ดังนี้

5.2.1 ข้อสอบที่เกิดจากการทำหน้าที่ต่างกันของข้อสอบและแบบสอบตามโครงการประเมินผลนักเรียนนานาชาติ ปี 2009

จากผลการวิจัย พบว่า ผลการทำหน้าที่ต่างกันของข้อสอบ (DIF) ที่มีการตรวจให้คะแนนทวิวิภาคและพหุวิภาค และการทำหน้าที่ต่างกันของแบบทดสอบ (DTF) ในแบบทดสอบการประเมินการรู้เรื่องวิทยาศาสตร์ระหว่างโมเดลที่ไม่ได้พิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ (โมเดลที่ 1) กับโมเดลที่มีการพิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบโดยการตัดข้อสอบที่มีการทำหน้าที่ต่างกันของข้อสอบและการทำหน้าที่ต่างกันเป็นขั้นที่แตกต่างกัน (โมเดลที่ 2) มีความแตกต่างกัน สามารถอภิปรายรายละเอียดได้ดังต่อไป

5.2.1.1 การศึกษาการทำหน้าที่ต่างกันของข้อสอบตามโครงการประเมินผลนักเรียนนานาชาติ ปี 2009

โครงการประเมินผลนักเรียนนานาชาติ (PISA = programme for international student assessment) เป็นการทดสอบทางการศึกษาระดับนานาชาติซึ่งถูกพัฒนาจากองค์กรที่เรียกว่า องค์การเพื่อความร่วมมือทางเศรษฐกิจและการพัฒนา (OECD = the organization for economic cooperation and development) การทดสอบเกิดขึ้นครั้งแรกเมื่อปี ค.ศ. 2000 และได้ดำเนินการทดสอบทุกๆ 3 ปี โดยมีเนื้อหาการทดสอบแบ่งเป็น 3 ด้าน ได้แก่ การรู้เรื่องทางวิทยาศาสตร์ (science literacy) การรู้เรื่องการอ่าน (reading literacy) และการรู้เรื่องคณิตศาสตร์ (mathematics literacy) ในส่วนของประเทศไทย โครงการประเมินผลนักเรียนนานาชาติ (PISA) มีศูนย์ดำเนินงานอยู่ที่สถานบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (สสทว.) ข้อมูลการประเมินผลมีประโยชน์ต่อองค์กรที่เกี่ยวข้อง เช่น กระทรวงศึกษาธิการ โรงเรียนมัธยมศึกษาทั้งสายสามัญและสายอาชีวศึกษา และสำนักประสานและพัฒนาการจัดการศึกษาท้องถิ่น ทั้งนี้ผลการประเมินจาก PISA ยังเป็นแหล่งข้อมูลที่สำคัญในระดับนโยบายและการดำเนินการติดตามแนวโน้มทั้งทางด้านคุณภาพการศึกษาและสิ่งแวดล้อมทางการศึกษา เพื่อการแก้ไขจุดบกพร่องและรักษาจุดแข็งในระบบการศึกษาของแต่ละองค์กร ดังเช่นรายงานผลการประเมิน PISA 2009 จากสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (สสวท.) (2553, 2554) ได้มีการรายงานผลการประเมินการรู้เรื่องการอ่าน วิทยาศาสตร์ และการรู้เรื่องคณิตศาสตร์ ของนักเรียนไทย พบว่า ผลการประเมินการรู้เรื่องการอ่านของนักเรียนไทยต่ำกว่าค่าเฉลี่ย OECD โดยมีคะแนนเฉลี่ย 421 อยู่ในช่วงอันดับประมาณ 50 ส่วนผลการประเมินการรู้เรื่องคณิตศาสตร์ต่ำกว่าค่าเฉลี่ย OECD โดยมีคะแนนเฉลี่ย 419 อยู่ที่ตำแหน่งอันดับประมาณ 52 และผลการประเมินการรู้เรื่องวิทยาศาสตร์ต่ำกว่าค่าเฉลี่ย OECD โดยมีคะแนนเฉลี่ย 425 อยู่ที่ตำแหน่ง

อันดับประมาณ 49 จากทั้งหมด 65 ประเทศ จากผลการประเมินดังกล่าวให้ทราบว่าคะแนนเฉลี่ยของนักเรียนไทยจากทั้ง 3 วิชามีคะแนนเฉลี่ยอยู่ในระดับที่ต่ำกว่าค่าเฉลี่ย ซึ่งสะท้อนระบบการจัดการศึกษาของประเทศจะต้องมีการพัฒนาระบบการเรียนการสอนให้สอดคล้องกับเนื้อหาของของข้อสอบ PISA เน้นให้นักเรียนรู้จักการวิเคราะห์และแก้ปัญหาในชีวิตได้อย่างแท้จริง นอกจากนี้เนื้อหาที่จัดการเรียนการสอนในหลักสูตรปัจจุบันควรที่จะมีความเชื่อมโยงไปยังข้อสอบ PISA ดังที่สัมพันธ์ พันธุ์พฤกษ์ (2557) ได้กล่าวถึงการนำแนวข้อสอบ PISA Liked มาใช้ในการจัดสอบวัดสมรรถนะของผู้เรียนในการทดสอบทางการศึกษาแห่งชาติขั้นพื้นฐาน หรือ O-Net โดยเน้นที่ไปที่ความเชื่อมโยงระหว่างรูปแบบข้อสอบเนื้อหาแต่ละวิชาในชีวิตประจำวัน นักเรียนจะต้องรู้จักคิดวิเคราะห์ ประมวลผล ซึ่งกระบวนการทั้งหมดจะต้องเกิดจากการเรียนรู้ในชั้นเรียน อีกทั้งครูต้องมีการปรับรูปแบบการเรียนการสอนเพื่อกระตุ้นให้นักเรียนรู้จักคิด สร้างสถานการณ์ในชั้นเรียนให้นักเรียนได้ลองเรียนรู้เพื่อฝึกฝน ไม่ใช่สอนให้เด็กท่องจำแต่เพียงอย่างเดียวเท่านั้น

ในส่วนการศึกษาการทำหน้าที่ต่างกันของข้อสอบ (DIF) ตามโครงการประเมินผลนักเรียนนานาชาติ หรือ PISA ปี ค.ศ. 2009 การประเมินการรู้เรื่องทางวิทยาศาสตร์ (science literacy) เป็นข้อสอบที่มีรูปแบบข้อสอบและเนื้อหาในการออกข้อสอบครอบคลุม 3 สมรรถนะหลักทางวิทยาศาสตร์ ได้แก่ การระบุประเด็นวิทยาศาสตร์ (identifying scientific issues) การอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ (explain phenomena scientifically) และการใช้ประจักษ์พยานทางวิทยาศาสตร์ (using scientific evidence) และสามารถจำแนกเป็นความรู้ทางวิทยาศาสตร์ 2 ส่วน ได้แก่ ส่วนที่หนึ่งความรู้วิทยาศาสตร์ ประกอบด้วย ระบบโลกและอวกาศ ระบบการดำรงชีวิต ระบบทางกายภาพ ระบบเทคโนโลยี และวิทยาศาสตร์และเทคโนโลยี และส่วนที่สองความรู้เกี่ยวกับวิทยาศาสตร์ ประกอบด้วย การอธิบายเชิงวิทยาศาสตร์ และการแสวงหาความรู้เชิงวิทยาศาสตร์ ในการวิจัยครั้งนี้ผลที่เกิดขึ้นจากข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันส่วนหนึ่งเป็นผลมาจากรูปแบบข้อสอบและเนื้อหาที่ใช้ในการทดสอบซึ่งส่วนหนึ่งเป็นข้อสอบที่ได้มีการทดสอบไปเมื่อปี 2006 เนื่องจาก ปี 2009 เนื้อหาข้อสอบจะเน้นการรู้เรื่องการอ่าน (reading literacy) เป็นวิชาหลักมีน้ำหนัก 60% ส่วนการรู้เรื่องทางวิทยาศาสตร์ (science literacy) และการรู้เรื่องคณิตศาสตร์ (mathematics literacy) ซึ่งเป็นวิชารองมีน้ำหนักอย่างละ 20% เท่ากัน ดังนั้นข้อสอบจึงเป็นฐานข้อสอบเดิมทั้งรูปแบบข้อสอบและเนื้อหายังคงเป็นข้อสอบที่มีลักษณะเดียวกับปี 2006 ซึ่งข้อสอบจากโครงการประเมินผลนักเรียนนานาชาตินั้นจะมีรอบระยะเวลาที่เปลี่ยนแปลงข้อสอบในแต่ละส่วนการรู้เรื่องที่แตกต่างกัน ทั้งนี้ในปี 2015 วิชาหลักจะเป็นการรู้เรื่องวิทยาศาสตร์ ซึ่งข้อสอบการรู้เรื่องวิทยาศาสตร์จะเป็นฐานข้อสอบชุดใหม่ ดังนั้นผลจากการทำหน้าที่ต่างกันของข้อสอบที่เกิดขึ้นในปี 2009 จึงสอดคล้องแนวคิดของ Park และ Lautenschlager (1990 อ้างถึงใน กาญจนา วัธนสุนทร, 2537) ซึ่งให้ข้อสังเกตเกี่ยวกับความลำเอียงหรือการทำหน้าที่

ต่างกันของข้อสอบมาจาก 2 ประเด็น คือ 1) อิทธิพลจากแหล่งความแปรปรวน จากข้อสอบและแหล่งความแปรปรวนจากผู้สอบ และ 2) แหล่งความแปรปรวนภายนอกดังกล่าวมีผลต่อการตอบข้อสอบในลักษณะที่ทำให้กลุ่มผู้สอบย่อยบางกลุ่มแตกต่างจากกลุ่มอื่นอย่างเป็นระบบทำให้เกิดการไม่ยุติธรรม นอกจากนี้ยังสอดคล้องกับกับการวิจัยของ Le (2009) ได้ศึกษาการทำหน้าที่ต่างกันของข้อสอบในแบบสอบการประเมินผลนักเรียนนานาชาติ (PISA) ในเนื้อหาประเมินการรู้เรื่องทางวิทยาศาสตร์ (science literacy) พบว่า ข้อสอบทำหน้าที่ต่างกันซึ่งส่วนหนึ่งมาจากรูปแบบข้อสอบและเนื้อหาที่ใช้ออกข้อสอบ และการศึกษาของ Wetzel และ Carstensen (2013) พบว่าข้อคำถามในข้อสอบการประเมินการรู้เรื่องการอ่านและการรู้เรื่องวิทยาศาสตร์ในที่ใช้ในข้อสอบปี 2009 เมื่อเปรียบเทียบกับข้อสอบในปี 2000 ประโยคคำถามมีความแตกต่างกันน้อย อีกทั้งคะแนนความสามารถอาจจะเองเอียงไปยังด้านใดด้านหนึ่งขึ้นอยู่กับว่าข้อสอบข้อนั้นได้รับอิทธิพลจากความยากตามเนื้อหาของข้อสอบ เช่นเดียวกันการศึกษาของ Adam (2003) และ Adam และคณะ (2007) ที่ได้มีการระบุถึงความคลาดเคลื่อนที่เกิดจากรูปแบบข้อสอบหรือการจัดพิมพ์ที่ผิดพลาดของข้อสอบฉบับย่อยในบางประเทศทำให้เกิดข้อสอบทำหน้าที่ต่างกันได้

นอกจากนี้ผู้วิจัยยังได้ศึกษาผลของข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันจากเกิดจากตัวแปรของนักเรียนที่แตกต่างกัน 5 กลุ่ม คือ เพศของนักเรียน (หญิง, ชาย) การใช้เวลาเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา (นักเรียนที่เรียนพิเศษ, นักเรียนที่ไม่เรียนพิเศษ) เศรษฐานะของครอบครัว (นักเรียนที่มีเศรษฐานะของครอบครัวสูง, นักเรียนที่มีเศรษฐานะของครอบครัวต่ำ) แหล่งทรัพยากรการศึกษาที่บ้าน (นักเรียนที่มีแหล่งทรัพยากรที่บ้านสูง, นักเรียนที่มีแหล่งทรัพยากรที่บ้านต่ำ) และความมั่งคั่งของครอบครัว (นักเรียนที่มีความมั่งคั่งของครอบครัวสูง, นักเรียนที่มีความมั่งคั่งของครอบครัวต่ำ) พบว่า ข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันเอนเอียงเข้าข้างนักเรียนหญิงมากกว่านักเรียนชาย เอนเอียงเข้าข้างนักเรียนที่เรียนพิเศษวิทยาศาสตร์นอกสถานศึกษามากกว่านักเรียนที่ไม่เรียนพิเศษนอกสถานศึกษา เอนเอียงเข้าข้างนักเรียนนักเรียนที่มีเศรษฐานะของครอบครัวต่ำมากกว่านักเรียนที่มีเศรษฐานะของครอบครัวสูง เอนเอียงเข้าข้างนักเรียนที่มีทรัพยากรของครอบครัวสูงมากกว่านักเรียนที่มีทรัพยากรของครอบครัวต่ำ และเอนเอียงเข้าข้างนักเรียนที่มีความมั่งคั่งทางครอบครัวสูงมากกว่านักเรียนที่มีความมั่งคั่งทางครอบครัวต่ำ สอดคล้องกับผลการวิจัยของ รักชนก ยี่สุ่นศรี (2544) ซึ่งได้วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบด้วยกระบวนการ ดีเอฟ ไอ ที พบว่า ข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันตามเพศทั้งในวิชาอังกฤษและคณิตศาสตร์ และการศึกษาของ Tuksino, Kanjanawasee, และ Pasiphol (2010) ซึ่งได้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ใช้ในการประเมินผลนักเรียนนานาชาติ (PISA) ปี 2006 ด้านการรู้เรื่องวิทยาศาสตร์ พบว่า ข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันในตัวแปรการเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษาและความมั่งคั่งของครอบครัว โดยที่นักเรียนที่เรียน

พิเศษวิทยาศาสตร์นอกสถานศึกษามีโอกาสทำข้อสอบได้ถูกมากกว่านักเรียนที่ไม่เรียนพิเศษนอกสถานศึกษา และนักเรียนที่มีความมั่งคั่งทางครอบครัวสูงมีโอกาสทำข้อสอบได้ถูกมากกว่านักเรียนที่มีความมั่งคั่งทางครอบครัวต่ำ นอกจากการศึกษาตามบริบทของประเทศไทย นักวิชาการจากต่างประเทศซึ่งได้ทำการศึกษาการทำหน้าที่ต่างกันของข้อสอบ พบว่า งานวิจัยครั้งนี้สอดคล้องจากการศึกษาของ Barnes และคณะ (2009) และ Bao และคณะ (2009) พบว่า ข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันตามเพศทำให้มีผลต่อการเองเอียงในการเลือกตอบข้อสอบ และการศึกษาของ Mahuteau และ Mavromaras (2009) พบว่า ข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันตามเศรษฐานะของครอบครัว โดยที่เศรษฐานะของครอบครัวนักเรียนมีผลคะแนนการประเมินผลโครงการประเมินผลนักเรียนนานาชาติ (PISA) ปี 2009 ซึ่งส่งผลต่อการเลือกเรียนของนักเรียนที่เข้าศึกษาต่อทั้งจากโรงเรียนรัฐและรัฐบาล

นอกจากข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันจากตัวแปรของนักเรียน ผู้วิจัยได้ยังได้ศึกษาโดยแยกตามสมรรถนะทางด้านวิทยาศาสตร์ 3 ด้านได้แก่ การระบุประเด็นทางวิทยาศาสตร์ การอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ และการใช้ประจักษ์พยานทางวิทยาศาสตร์ พบว่า **การระบุประเด็นทางวิทยาศาสตร์**เอนเอียงเข้าข้างนักเรียนหญิงมากกว่านักเรียนชาย เอนเอียงเข้าข้างนักเรียนที่ไม่เรียนพิเศษมากกว่านักเรียนที่เรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา เอนเอียงเข้าข้างนักเรียนที่มีเศรษฐานะต่ำมากกว่านักเรียนที่มีเศรษฐานะสูง เอนเอียงเข้าข้างนักเรียนที่มีทรัพยากรครอบครัวต่ำมากกว่านักเรียนที่มีทรัพยากรครอบครัวสูง และเอนเอียงเข้าข้างนักเรียนที่มีความมั่งคั่งครอบครัวต่ำมากกว่านักเรียนที่มีความมั่งคั่งครอบครัวสูง **การอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์**เอนเอียงเข้าข้างนักเรียนชายมากกว่านักเรียนหญิง เอนเอียงเข้าข้างนักเรียนที่เรียนพิเศษมากกว่านักเรียนที่ไม่เรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา เอนเอียงเข้าข้างนักเรียนที่มีเศรษฐานะต่ำมากกว่านักเรียนที่มีเศรษฐานะสูง เอนเอียงเข้าข้างนักเรียนที่มีทรัพยากรครอบครัวสูงมากกว่านักเรียนที่มีทรัพยากรครอบครัวต่ำ และเอนเอียงเข้าข้างนักเรียนที่มีความมั่งคั่งครอบครัวต่ำมากกว่านักเรียนที่มีความมั่งคั่งครอบครัวสูง ส่วน**การใช้ประจักษ์พยานทางวิทยาศาสตร์**ไม่เอนเอียงเข้าข้างในตัวแปรเพศและไม่เอนเอียงเข้าข้างในตัวแปรนักเรียนที่เรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา แต่เอนเอียงเข้าข้างนักเรียนที่มีเศรษฐานะสูงมากกว่านักเรียนที่มีเศรษฐานะต่ำ เอนเอียงเข้าข้างนักเรียนที่มีทรัพยากรครอบครัวสูงมากกว่านักเรียนที่มีทรัพยากรครอบครัวต่ำ และเอนเอียงเข้าข้างนักเรียนที่มีความมั่งคั่งครอบครัวสูงมากกว่านักเรียนที่มีความมั่งคั่งครอบครัวต่ำ ซึ่งผลการศึกษาการทำหน้าที่ต่างกันของข้อสอบแยกตามสมรรถนะทางด้านวิทยาศาสตร์ สอดคล้องกับผลการศึกษาของ Le (2006) ได้วิเคราะห์การทำหน้าที่ต่างกันของสอบตามสมรรถนะการรู้เรื่องวิทยาศาสตร์ พบว่า ข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันตามเพศ และแยกตามสมรรถนะทางวิทยาศาสตร์ 2 ด้าน คือการระบุประเด็นทางวิทยาศาสตร์ข้อสอบส่วนใหญ่เอนเอียงเข้าข้างนักเรียน

หญิงมากกว่านักเรียนชาย และการอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ข้อสอบส่วนใหญ่เอนเอียงเข้าข้างนักเรียนชายมากกว่านักเรียนหญิง ส่วนการใช้ประจักษ์พยานทางวิทยาศาสตร์ไม่เอนเอียงเข้าข้างในตัวแปรเพศ

อีกประเด็นที่น่าสนใจก็คือ ในการวิจัยครั้งนี้สำหรับข้อสอบที่มีการให้คะแนนพหุวิภาค ผู้วิจัยได้ใช้การศึกษาการทำหน้าที่ต่างกันของข้อสอบโดยพิจารณาค่าสถิติตัวประมาณค่าอิทธิพลการทำหน้าที่ต่างกันของข้อสอบลัว-อเกรสตีปกติกับอัตราส่วนแต้มต่อ (standardized Liu-Agresti common log-odds ratio DIF effect estimator) ร่วมกับการศึกษาการทำหน้าที่ต่างกันเป็นขั้น (differential step functioning: DSF) จากขนาดอิทธิพลการทำหน้าที่ต่างกันเป็นขั้น 3 ระดับ คือ ขนาดเล็ก ขนาดกลาง และขนาดใหญ่ โดยใช้สถิติตัวประมาณค่าอัตราส่วนลอกเลขออกร่วมระดับขั้นสะสม (the cumulative step-level log-odds ratio estimator) ซึ่งทำให้เพิ่มสารสนเทศในการศึกษาการทำหน้าที่ต่างกันของข้อสอบตามการระบุขนาดอิทธิผลในแต่ละขั้นของการให้คะแนนในข้อสอบที่มีการให้คะแนนพหุวิภาค Penfield (2007b, 2008, 2010a) ได้กล่าวถึงความแตกต่างระหว่างกลุ่มในแต่ละขั้น (step) ของข้อสอบที่มีการตรวจให้คะแนนพหุวิภาคแต่ละขั้นแสดงถึงการมีโอกาสของการได้เปรียบ (probability of advancing) หรือ stepping และแต่ละขั้นอาจเกิดเหตุการณ์ที่กลุ่มอ้างอิงได้เปรียบกลุ่มเปรียบเทียบ หรือกลุ่มเปรียบเทียบได้เปรียบกลุ่มอ้างอิงก็ได้ ส่วน Padilla และคณะ (2011) ดำเนินการศึกษาการทำหน้าที่ต่างกันของข้อสอบในข้อสอบที่มีการตรวจให้คะแนนพหุวิภาคตามโครงการประเมินผลนักเรียนนานาชาติ (PISA) พิจารณาจากขนาดอิทธิพลขนาดกลาง และขนาดใหญ่ เป็นขนาดอิทธิพลที่ทำให้เกิดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน นอกจากนี้ Penfield, Gattamorta, และ Childs (2009) ได้ให้ข้อเสนอแนะเพิ่มเติมว่าข้อสอบทำหน้าที่ต่างกันมาจากผลในระดับข้อสอบ (item-level effects) หรือผลในระดับคะแนน (specific score levels) มักจะส่งผลไปยังการเกิดการทำหน้าที่ต่างกันของข้อสอบได้ ทั้งนี้แนวคิดการศึกษาการทำหน้าที่ต่างกันของข้อสอบสำหรับข้อสอบที่มีการตรวจให้คะแนนพหุวิภาคร่วมกับการทำหน้าที่ต่างกันเป็นขั้นเป็นแนวคิดใหม่ทางด้านการศึกษาการทำหน้าที่ต่างกันของข้อสอบ เป็นแนวคิดที่พยายามนำเสนอค่าสารสนเทศในระดับขั้น (Step) ในการพิจารณาการทำหน้าที่ต่างกันเป็นขั้น (DSF) ในข้อสอบที่มีการตรวจให้คะแนนพหุวิภาคโดยใช้ขนาดอิทธิผลเข้ามาศึกษาประกอบการพิจารณา ซึ่งแนวคิดแบบดั้งเดิมนั้นการระบุความลำเอียงหรือการกำจัดแหล่งความลำเอียงจากทั้งข้อสอบและแบบสอบจำเป็นอย่างยิ่งที่จะต้องบอกได้ว่าข้อสอบในแบบสอบนั้นเป็นข้อสอบที่มีความยุติธรรมหรือไม่ โดยที่การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการตรวจให้คะแนนพหุวิภาคมักระบุความลำเอียงในระดับกลุ่มข้อสอบหรือลักษณะเกณฑ์การให้คะแนนน้อยมาก โดยเฉพาะข้อสอบที่มีลักษณะแบบเลือกตอบ (multiple-choice items) ทำให้เกิดความผิดพลาดในการตรวจสอบความลำเอียง เพราะความลำเอียงที่เกิดขึ้นถูกพิจารณาจากคะแนนเป็นระดับเดียวเท่านั้น

ดังนั้นการนำเทคนิควิธีการทำหน้าที่ต่างกันเป็นขั้น (DSF) จะช่วยให้ผู้วิจัยสามารถระบุความลำเอียงที่อาจจะเกิดขึ้นในระดับการให้คะแนน (score levels) ในระดับต่างๆ เพื่อนำเสนอค่าสารสนเทศประกอบการพิจารณาความลำเอียงในข้อสอบที่มีการตรวจให้คะแนนพหุวิภาค

5.2.1.2 การศึกษาการทำหน้าที่ต่างกันของแบบสอบตามโครงการประเมินผลนักเรียนนานาชาติ ปี 2009

ในการวิจัยครั้งนี้ ผู้วิจัยได้ศึกษาผลของแบบสอบที่ตรวจพบการทำหน้าที่ต่างกันจากตัวแปรของนักเรียน 5 ตัวแปร คือ เพศของนักเรียน (หญิง, ชาย) การใช้เวลาเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา (นักเรียนที่เรียนพิเศษ, นักเรียนที่ไม่เรียนพิเศษ) เศรษฐานะของครอบครัว (นักเรียนที่มีเศรษฐานะของครอบครัวสูง, นักเรียนที่มีเศรษฐานะของครอบครัวต่ำ) แหล่งทรัพยากรการศึกษาที่บ้าน (นักเรียนที่มีแหล่งทรัพยากรที่บ้านสูง, นักเรียนที่มีแหล่งทรัพยากรที่บ้านต่ำ) และความมั่งคั่งของครอบครัว (นักเรียนที่มีความมั่งคั่งของครอบครัวสูง, นักเรียนที่มีความมั่งคั่งของครอบครัวต่ำ) จากผลการศึกษาการทำหน้าที่ต่างกันของแบบสอบ โดยการตรวจสอบความแปรปรวนอิทธิพลการทำหน้าที่ต่างกันของข้อสอบ (the DIF effect variance) แบ่งเป็น 3 ระดับด้วยกัน คือ อิทธิพลขนาดเล็ก (S=small DIF effect variance) อิทธิพลขนาดกลาง (M=medium DIF effect variance) และอิทธิพลขนาดใหญ่ (L=large DIF effect variance) ทั้งนี้เมื่อผู้วิจัยใช้ระดับในการศึกษาอิทธิพลขนาดกลาง (M) และขนาดใหญ่ (L) ซึ่งเป็นเกณฑ์ในการศึกษาแบบสอบที่ตรวจพบการทำหน้าที่ต่างกัน ในตัวแปรเศรษฐานะของครอบครัว พบว่า แบบสอบที่ตรวจพบการทำหน้าที่ต่างกันมากที่สุด 4 ฉบับ คิดเป็นร้อยละ 44.44 รองลงมาคือความมั่งคั่งของครอบครัว พบว่า แบบสอบที่ตรวจพบการทำหน้าที่ต่างกัน 3 ฉบับ คิดเป็นร้อยละ 33.33 และแหล่งทรัพยากรการศึกษาที่บ้าน พบว่า แบบสอบที่ตรวจพบการทำหน้าที่ต่างกัน 2 ฉบับ คิดเป็น ร้อยละ 22.22 ส่วนเพศและการใช้เวลาเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษาของนักเรียนไม่พบแบบสอบที่ตรวจพบการทำหน้าที่ต่างกัน ซึ่งจากผลการวิจัยดังกล่าวทำให้ทราบว่าข้อสอบ PISA ปี 2009 ความลำเอียงที่เกิดขึ้นไม่ได้มีเฉพาะในระดับข้อสอบเท่านั้น ยังมีความลำเอียงที่เกิดขึ้นในระดับแบบสอบอีกด้วย โดยพบแบบสอบที่ตรวจพบการทำหน้าที่ต่างกันมากที่สุดในตัวแปรเศรษฐานะของครอบครัว ส่วนตัวแปรเพศและตัวแปรนักเรียนที่เรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา พบว่า มีข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันแต่เมื่อนำมาศึกษาการทำหน้าที่ต่างกันของแบบสอบไม่พบแบบสอบที่ตรวจพบการทำหน้าที่ต่างกัน ซึ่งผลที่ได้จากการวิจัยครั้งนี้ทำให้ได้ข้อค้นพบที่เป็นประโยชน์ในการศึกษาความยุติธรรมของข้อสอบและแบบสอบว่าการจะตัดข้อสอบข้อที่ตรวจพบว่ามีความลำเอียงนั้นสิ่งที่ต้องคำนึงที่นอกเหนือจาก ด้วยความเที่ยง

(reliability) และความตรง (validity) ที่ได้มาตรฐาน (Holland & Wainer, 1993; Embretson & Reise, 2000; Kunnan, 2000; de Ayala, 2008; Demar, 2010; ศิริชัย กาญจนวาสี, 2550ก) การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและแบบสอบก็มีความสำคัญเช่นเดียวกัน การเลือกตัดข้อสอบที่พบความลำเอียงควรจะมีการตรวจสอบทั้งฉบับด้วยว่าเมื่อตัดข้อสอบข้อนั้นออกไปแล้วแบบสอบยังตรวจพบว่ามีความลำเอียงหรือไม่ เพราะบางครั้งการตัดข้อสอบที่ตรวจพบความลำเอียงออกไปทั้งหมดอาจจะไม่ทำให้ความลำเอียงในแบบสอบลดลงไป สอดคล้องกับการวิจัยของ Cho และคณะ (2010) ทำการศึกษามาตรวัดพฤติกรรมต่อต้านสังคมของเด็กวัยรุ่นและวัยผู้ใหญ่ตอนต้นจากทั้ง 2 กลุ่มคือกลุ่มที่ถูกเลี้ยงมาจากทั้งพ่อและแม่กับกลุ่มที่ถูกเลี้ยงมาจากพ่อแม่เลี้ยงเดี่ยว พบว่า ทั้งสองกลุ่มที่ทำการทดสอบมาตรวัดวัดพฤติกรรมต่อต้านสังคมของเด็กวัยรุ่นและวัยผู้ใหญ่ตอนต้น พบว่า มีนัยสำคัญทางสถิติ แต่เมื่อทำการศึกษาการทำหน้าที่ต่างกันของแบบสอบแล้วทั้งสองกลุ่มไม่พบความแตกต่าง ส่วน Gong (2012) ได้กล่าวเพิ่มเติมว่า การตรวจสอบการทำหน้าที่ต่างกันของแบบสอบเหมือนกับกระบวนการทดสอบการทำหน้าที่ต่างกันของแบบสอบมักใช้ทดสอบภาพรวมของข้อสอบมากกว่าเป็นการทดสอบระดับข้อสอบ การทดสอบการทำหน้าที่ต่างกันของแบบสอบมีความสำคัญเนื่องจากส่งผลต่อการการตัดสินใจเลือกข้อสอบเป็นครั้งสุดท้ายตามพื้นฐานของสมรรถนะของนักเรียนที่ปรากฏ ข้อสอบอาจจะพบว่ามีความแตกต่างกันระหว่างกลุ่มที่ศึกษาแต่นั้นก็ไม่มีผลเชิงรูปธรรมเสมอไปที่จะบอกได้ว่าเป็นผลมาจากแบบสอบทั้งฉบับ

อีกประเด็นที่น่าสนใจ ก็คือ การตัดข้อสอบบางข้อที่ตรวจพบการทำหน้าที่ต่างกันออกไปเป็นการลดแหล่งความคลาดเคลื่อนที่เกิดจากตัวแปรของนักเรียนที่มีความแตกต่างกัน สอดคล้องกับเหตุผลของการตัดข้อสอบบางข้อออกไปจากงานวิจัยของ Tuksino, Kanjanawasee, และ Pasiphol (2010) และ Sirikit, Pasiphol, และ Kanjanawasee (2012) ซึ่งให้เหตุผลการตัดข้อสอบออกไปบางข้อเพื่อการลดความคลาดเคลื่อนที่อาจจะเกิดขึ้นจากตัวแปรที่นำมาศึกษา เมื่อพิจารณาจากเกณฑ์ที่นำมาตัดข้อสอบ คือ 1) ข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันเนื่องจากตัวแปรนักเรียนมากกว่าหนึ่งตัวแปร 2) ตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันบางข้อออกไปแล้วไม่ควรเกินร้อยละ 20 ของข้อสอบที่มีอยู่ในแต่ละฉบับ โดยเงื่อนไขการตัดข้อสอบสำหรับข้อสอบที่มีการให้คะแนนทวิวิภาคพิจารณาค่าสถิติแมนเทล-แฮนส์เซลร่วมกับอัตราส่วนแต้มต่อมาตรฐาน (standardized mantel-haenszel log-odds ratio) ส่วนข้อสอบที่มีการให้คะแนนพหุวิภาคพิจารณาค่าสถิติตัวประมาณค่าอิทธิพลการทำหน้าที่ต่างกันของข้อสอบลัว-อเกรสตีปกติกับอัตราส่วนแต้มต่อ (standardized Liu-Agresti common log-odds ratio DIF effect estimator) ร่วมกับขนาดอิทธิพลการทำหน้าที่ต่างกันเป็นขั้น ส่วนการทำหน้าที่ต่างกันของแบบสอบพิจารณาจากขนาดอิทธิพลการทำหน้าที่ต่างกันของแบบสอบ ขนาดกลาง (M) และขนาดใหญ่ (L) และ 3) ข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันที่ถูกตัดไปเมื่อพิจารณาการทำหน้าที่ต่างกันของแบบสอบควรทำให้

อิทธิพลแบบสอบทำหน้าที่ต่างกันลดลงหรือเท่าเดิมได้ซึ่งต้องตรวจสอบความตรงเชิงโครงสร้างแบบสอบภายหลังข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันออกไป ซึ่งผู้วิจัยได้คำนึงความเที่ยง (reliability) ของแบบสอบทั้งก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า ค่าความเที่ยงมีค่าลดลงหลังตัดข้อสอบบางข้อที่ตรวจพบว่ามีการทำหน้าที่ต่างกันออกไป จากผลการวิจัยที่เกิดขึ้นสอดคล้องกับผลการวิจัยของ รักชนก ยี่สุ่นศรี (2544) ซึ่งพบว่า ถ้าตัดข้อสอบออกจากแบบสอบจำนวนมากขึ้นจะทำให้ค่าความเที่ยงของแบบสอบมีค่าลดดง ความเที่ยงที่ลดลงจากการตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน อันเป็นผลมาจากความแปรปรวนของคะแนนที่สังเกตได้มีค่าลดลงหรือความแปรปรวนของคะแนนความคลาดเคลื่อนมีค่าสูงขึ้น เพราะความเที่ยง ก็คือ อัตราส่วนระหว่างความแปรปรวนจากคะแนนจริงเทียบกับความแปรปรวนของคะแนนที่วัดได้การตัดข้อสอบย่อมส่งผลต่อค่าความเที่ยงที่ตามมา (ศิริชัย กาญจนวาสี, 2544; นงลักษณ์ วิรัชชัย, 2542) ทั้งนี้ความแปรปรวนมีทั้งความแปรปรวนของส่วนที่ตรงประเด็นตามลักษณะที่มุ่งวัดกับส่วนที่ไม่ตรงประเด็น ดังนั้นสัดส่วนความแปรปรวนที่เที่ยงลดลงนั้นอาจมาจากการลดลงของความแปรปรวนในส่วนที่ไม่ตรงประเด็นตามลักษณะที่มุ่งวัด ทำให้แบบสอบมีค่าความเที่ยงลดลงแต่ความตรงสูงขึ้น นอกจากนี้เพื่อเป็นการตรวจสอบความตรงให้สอดคล้องกับผู้เชี่ยวชาญได้กล่าวไว้ ผู้วิจัยจึงได้ตรวจสอบความตรงเชิงโครงสร้างโดยการวิเคราะห์องค์ประกอบเชิงยืนยันของแบบสอบภายหลังที่ตัดข้อสอบที่พบการทำหน้าที่ต่างกันออกไป ทั้งนี้จากผลการวิจัย ในส่วนแบบสอบรวมทั้งหมดและแบบสอบแยก 9 ฉบับภายหลังที่มีการตัดข้อสอบตามเกณฑ์ที่กำหนด พบว่า แบบสอบที่ตัดข้อสอบทำหน้าที่ต่างกันออกไปแล้วยังคงมีความเหมาะสมที่จะนำเฉพาะข้อสอบบางข้อที่ได้รับการคัดเลือกเข้ามาศึกษา ซึ่งข้อดีของการวิเคราะห์องค์ประกอบเชิงยืนยันเป็นประโยชน์ในการวิเคราะห์องค์ประกอบและพัฒนาสเกลองค์ประกอบ อีกทั้งยังสามารถเป็นเครื่องมือในการตรวจสอบความตรงเชิงโครงสร้างได้เช่นเดียวกัน (นงลักษณ์ วิรัชชัย, 2542)

5.2.1.3 ความสัมพันธ์ระหว่างการทำหน้าที่ต่างกันของข้อสอบและแบบสอบตามโครงการประเมินผลนักเรียนนานาชาติ ปี 2009

ในการวิจัยครั้งนี้ผู้วิจัยได้ดำเนินการศึกษาการทำหน้าที่ต่างกันของข้อสอบทั้งข้อสอบที่มีการตรวจให้คะแนนทวิวิภาคและพหุวิภาค จากผลการวิจัยก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า การทำหน้าที่ต่างกันของข้อสอบสามารถเชื่อมโยงการทำหน้าที่ต่างกันของแบบสอบในแบบสอบทั้ง 9 ฉบับ แยกพิจารณาตัวแปรทั้ง 5 โดยใช้เกณฑ์การศึกษาอิทธิพลขนาดกลาง (M) และขนาดใหญ่ (L) มาเกณฑ์ในการศึกษาแบบสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า ในการศึกษาครั้งนี้ขนาดอิทธิพลที่ปรากฏมักเป็นขนาดอิทธิพลขนาดกลาง ไม่ปรากฏขนาด

อิทธิพลขนาดใหญ่ ซึ่งอิทธิผลการทำหน้าที่ต่างกันของแบบสอบที่มีขนาดกลาง (M) ที่สอดคล้อง 2 กรณี คือ กรณีที่ 1 ไม่พบข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันทั้งการตรวจให้คะแนนทวิวิภาคแต่ตรวจพบในข้อสอบที่มีการตรวจให้คะแนนพหุวิภาค ซึ่งเกิดขึ้นในตัวแปรเศรษฐานะของครอบครัวจากแบบสอบฉบับที่ 3, 7, และ 10 ตัวแปรตัวทรัพยากรของครอบครัวจากแบบสอบฉบับที่ 3 ตัวแปรความมั่งคั่งของครอบครัวจากแบบสอบฉบับที่ 3, 12, และ 13 และกรณีที่สอง คือ พบข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันทั้งการตรวจให้คะแนนทวิวิภาคและพหุวิภาค เกิดขึ้นในตัวแปร เศรษฐานะของครอบครัวจากแบบสอบฉบับที่ 12 และตัวแปรแหล่งทรัพยากรของครอบครัวจากแบบสอบฉบับที่ 13 ดังนั้นจากผลการวิเคราะห์ดังกล่าวสอดคล้องกับข้อเสนอของ Penfield และ Algina (2006) ที่ได้กล่าวไว้ว่า การประมาณค่าผลการทำหน้าที่ต่างกันของข้อสอบโดยใช้ตัวแปรที่มีการแยกเป็นชั้น ภายในแบบสอบส่งผลกระทบโดยตรงการเกิดการทำหน้าที่ต่างกันของข้อสอบ การวัดการทำหน้าที่ต่างกันของแบบสอบสามารถให้ดัชนีผลการทำหน้าที่ต่างกันของข้อสอบซึ่งระดับการทำหน้าที่ต่างของแบบสอบขนาดใหญ่มีผลต่อการทำหน้าที่ต่างกันของข้อสอบมากกว่าการทำหน้าที่ต่างกันในแบบสอบขนาดเล็ก

5.2.2 ผลการเปรียบเทียบประสิทธิภาพโมเดลมูลค่าเพิ่มการจัดการศึกษาวิทยาศาสตร์

จากผลการวิจัย พบว่า ผลการเปรียบเทียบประสิทธิภาพโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์เมื่อใช้โมเดลมูลค่าเพิ่มพหุระดับของ 2 โมเดล พบว่า โมเดลที่ไม่ได้พิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ (โมเดลที่ 1) กับโมเดลที่มีการพิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบโดยการตัดข้อสอบที่มีการทำหน้าที่ต่างกันของข้อสอบและการทำหน้าที่ต่างกันเป็นขั้นที่แตกต่างกัน (โมเดลที่ 2) มีประสิทธิภาพที่แตกต่างกัน และโมเดลที่มีการพิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบโดยการตัดข้อสอบที่มีการทำหน้าที่ต่างกันของข้อสอบและการทำหน้าที่ต่างกันเป็นขั้นที่แตกต่างกันมีประสิทธิภาพสูงกว่าโมเดลที่ไม่ได้พิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ นอกจากนี้เมื่อศึกษาคะแนนมูลค่าเพิ่มกับ 3 ตัวแปร คือ เศรษฐานะของครอบครัว แหล่งทรัพยากรการศึกษาที่บ้าน และความมั่งคั่งของครอบครัว ทั้งก่อนและหลังตัดข้อสอบที่ทำหน้าที่ต่างกันมีความสัมพันธ์กัน และคะแนนมูลค่าเพิ่มกับทั้ง 3 ตัวแปรภายหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันออกไปแล้วในการวิเคราะห์มูลค่าเพิ่มมีขนาดลดลงเมื่อเปรียบเทียบกับการวิเคราะห์ที่ไม่ได้ตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันออกไป และผลการเปรียบเทียบความสอดคล้องของการจัดอันดับคุณภาพโรงเรียนจากคะแนนมูลค่าเพิ่มโดยสถิติทดสอบ Wilcoxon Signed Ranks พบว่า โมเดลที่ 1 และโมเดลที่ 2 มีการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่มไม่สอดคล้องกัน ส่วนผลการเปรียบเทียบความสอดคล้องของการจัดกลุ่มคุณภาพ

โรงเรียน 5 กลุ่ม ตามคะแนนมูลค่าเพิ่ม พบว่า โมเดลที่ 1 กับโมเดลที่ 2 มีการจัดกลุ่มคุณภาพโรงเรียนที่สอดคล้องกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 สามารถอภิปรายรายละเอียดได้ดังต่อไป

5.2.2.1 ความสำคัญของการนำโมเดลมูลค่ามาใช้ในการพัฒนาคุณภาพการจัดการศึกษาวิทยาศาสตร์

การวิเคราะห์มูลค่าเพิ่มหรือโมเดลมูลค่าเพิ่มเป็นวิธีการเปรียบเทียบความสามารถของโรงเรียนและครูที่มีความยุติธรรมที่สุด โดยแนวคิด “มูลค่าเพิ่ม” หรือ value-added Model (VAM) เป็นแนวคิดทางด้านเศรษฐศาสตร์ ที่มีประสิทธิภาพสูงสุดในการประเมินประสิทธิผลขององค์กร แนวคิดมูลค่าเพิ่มถูกนำเข้ามาใช้ทางการศึกษาเพื่อประเมินความสามารถในการดำเนินงานของโรงเรียน ในยุคศตวรรษที่ 20 ซึ่งมาจากพระราชบัญญัติการศึกษาของประเทศสหรัฐอเมริกา ที่เรียกกันว่า no child left behide act of 2001 (NCLB) โดยการสร้างมูลค่าเพิ่มในผลการเรียนรู้เป็นการเปรียบเทียบระหว่างคะแนนผลการเรียนรู้ที่เกิดขึ้นจริงหรือสังเกตได้ (observed scores) กับคะแนนผลการเรียนรู้ที่ทำนายได้ (predicted scores) ทั้งนี้ประสิทธิภาพของโมเดลมูลค่าเพิ่มยังช่วยขจัดความลำเอียงที่สามารถอธิบายการดำเนินงานของครูและโรงเรียนได้อย่างถูกต้องและแม่นยำ คะแนนมูลค่าเพิ่มยังมีประโยชน์ในการประเมินและจัดอันดับ (ranking) โดยการควบคุมตัวแปรของนักเรียนหรือโรงเรียน เช่น ภูมิศาสตร์ที่เกี่ยวข้องกับนักเรียน ลักษณะทางเศรษฐานะการศึกษาของครอบครัว พื้นฐานด้านภาษา และสิ่งแวดล้อมที่เกี่ยวข้องต่างๆ (ศิริชัย กาญจนวาสี, 2550ข; Manzi et al., 2010) นอกจากนี้แนวคิดมูลค่าเพิ่ม (value-added approach) ยังเป็นแนวคิดมีความโดดเด่นทางการประเมินสามารถนำมาใช้ประเมินครูในสถานศึกษาโดยพิจารณาจากคะแนนการทดสอบของนักเรียน ทั้งนี้ผลการประเมินที่ได้ยังสามารถนำไปใช้วัดและพัฒนาคุณภาพการสอนของครูในโรงเรียนได้อีกด้วย (Hanushek, 1971, Murnane, 1975, Rockoff, 2004, Rivkin et al., 2005, Aaroson et al., 2007, Kane and Staiger, 2008 cited in Chetty et al., 2011) ในสถาบันการศึกษาสามารถนำแนวคิดมูลค่าเพิ่มมาเป็นส่วนหนึ่งของระบบความรับผิดชอบต่อการศึกษา (educational accountability system) (American Statistical Association, 2014) หรือกำหนดเป็นนโยบายของสถานศึกษาผ่านผลการประเมินการเรียนรู้ของนักเรียนที่สามารถเชื่อมโยงไปยังคุณภาพของครูแต่ละคนในสถานศึกษา โดยคำนึงถึง บทบาท/อัตลักษณ์ของโรงเรียน บุคลาครูทุกๆ คน ทัศนคติของนักเรียน ครอบครัวหรือพื้นฐานทางบ้านต่อพัฒนาการเรียนรู้ของนักเรียน ดังเช่น งานวิจัยของ Newton และคณะ (2010) ได้วิจัยประสิทธิผลของครูตามโมเดลมูลค่าเพิ่มโดยการสำรวจจากโมเดลมูลค่าเพิ่มตามบริบทของสถานศึกษา ซึ่งจากการวิจัยชี้ให้เห็นว่า

ข้อดีของการนำแนวคิดมูลค่ามาใช้ในสถานศึกษา ได้แก่ 1) การกำหนดรายวิชาให้เหมาะสมระหว่างของครูกับนักเรียนสามารถพิจาณาได้จากจัดอันดับการสอนของครูในวิชาสอนที่แตกต่างกัน 2) จัดการสอนของครูให้เหมาะสมกับนักเรียนเพื่อช่วยลดปัญหาเรื่องการเรียนรู้ที่อาจจะเกิดขึ้นต่อการเรียนรู้ของนักเรียน ซึ่งครูบางท่านไม่สามารถสอนได้ตลอดปีการศึกษา 3) การกำหนดเป็นนโยบายของสถาบันการศึกษาโดยการใช้มูลค่าเพิ่มมาประเมินสถาบันหรือความรับผิดชอบของครูต่อคุณภาพของนักเรียน นอกจากนี้ โมเดลการประเมินด้วยโมเดลมูลค่าเพิ่ม (value-added assesment models: VAAs) ยังสามารถนำมาใช้ประกอบการตัดสินใจของสถาบันการศึกษา เช่น การกำหนดค่าตอบแทนบุคลากร การประเมินหรือจัดอันดับครู การจ้างหรือยกเลิกการจ้างงานของครู การให้รางวัลตอบแทน หรือแม้กระทั้งการยุบสถาบันการศึกษา

5.2.2.2 การคัดเลือกตัวแปรเพื่อนำมาศึกษาประสิทธิภาพโมเดลมูลค่าเพิ่มการจัดการศึกษาวิทยาศาสตร์

จากการตรวจสอบเงื่อนไขการวิเคราะห์พหุระดับเบื้องต้นโดยการตรวจสอบความสัมพันธ์ระหว่างตัวแปรจากการตรวจสอบภาวะร่วมเส้นตรงพหุ (multicollinearity) และการตรวจสอบความสอดคล้องกับสมการการวิเคราะห์ถดถอย พบว่า ผลการวิเคราะห์ค่าสหสัมพันธ์ของตัวแปรในระดับนักเรียนก่อนและหลังข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า ความสัมพันธ์กับตัวแปรทำนายทั้งหมดกับตัวแปรตามมีนัยสำคัญทางสถิติที่ระดับ .01 ทั้งก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน ส่วนผลการตรวจสอบภาวะร่วมเส้นตรงพหุ (multicollinearity) ก่อนและหลังตัดข้อสอบที่ทำหน้าที่ต่างกันในตัวแปรระดับนักเรียนและสถานศึกษา โดยการวิเคราะห์ถดถอยพหุ (multiple regression) โดยพิจารณาค่าความคลาดเคลื่อนยินยอม (Tolerance) และค่า VIF (variance inflation factor) ซึ่งค่าความคลาดเคลื่อนยินยอม (Tolerance) ควรสูงกว่า 0.1 และค่า VIF ควรมีค่าน้อยกว่า 10 (Hair et al., 1998) ส่วนค่าดัชนีเงื่อนไข (condition index) ไม่ควรเกิน 30 (สมถวิล วิจิตรวรรณา, สุภมาส อังศุโชติ และรัชนีกูล ภิญโญภานุวัฒน์, 2553) พบว่า ไม่เกิดปัญหาตัวแปรทำนายมีความสัมพันธ์กันสูงหรือไม่เกิดภาวะร่วมเส้นตรงพหุก่อนและหลังการตัดข้อสอบทำหน้าที่ต่างกันของตัวแปรทำนายทั้งหมด ซึ่งผลที่ได้มีค่าความคลาดเคลื่อนยินยอม (tolerance) สูงกว่า 0.1 และค่า VIF น้อยกว่า 10 ส่วนค่าดัชนีเงื่อนไข (condition index) ไม่เกิน 30 ดังนั้นจึงไม่ต้องตัดตัวแปรทำนายระดับนักเรียนและสถานศึกษาตัวใดทิ้ง เนื่องจากในการวิจัยครั้งนี้ผู้วิจัยได้ใช้ฐานข้อมูลจากโครงการประเมินผลนักเรียนนานาชาติ (PISA) 2009 ซึ่งมีตัวแปรทั้งในระดับนักเรียนและสถานศึกษาเป็นจำนวนมาก ไม่สามารถที่จะนำตัวแปรทุกตัวเข้ามาศึกษา ดังนั้นผู้วิจัยจึงใช้หลักการทดสอบทางสถิติ

เข้ามาช่วยทดสอบ ซึ่งก่อนการวิเคราะห์มูลค่าเพิ่มผู้วิจัยได้ใช้การตรวจสอบเงื่อนไขเบื้องต้นสำหรับตัวแปรทั้งในระดับนักเรียนและสถานศึกษาด้วยการตรวจสอบภาวะร่วมเส้นตรงพหุ (multicollinearity) ซึ่งทั้งแปรพยากรณ์ต้องเป็นอิสระต่อกันหรือไม่มีความสัมพันธ์กันเอง ถ้าตัวแปรพยากรณ์มีความสัมพันธ์กันเองสูงมากจะทำให้เพิ่มโอกาสของความคลาดเคลื่อนประเภทที่ 2 (Type II error) มากขึ้น ดังนั้น ผลการวิจัยดังกล่าวจึงมีความถูกต้องและเป็นไปตามหลักวิชาการในการนำตัวแปรที่ผ่านการคัดเลือกพร้อมที่จะนำมาวิเคราะห์มูลค่าเพิ่มพหุระดับ (นงลักษณ์ วิรัชชัย, 2542; สมถวิล วิจิตรวรรณา, สุภมาส อังศุโชติ, และรัชนีกูล ภิญโญภานวัฒน์ ,2553 อ้างถึงใน บุญเรียง ขจรศิลป์ , 2555)

5.2.2.3 ผลการเปรียบเทียบประสิทธิภาพโมเดลมูลค่าเพิ่มพหุระดับเพื่อการประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์

ก่อนการเปรียบเทียบประสิทธิภาพโมเดลมูลค่าเพิ่มพหุระดับ ผู้วิจัยได้ตรวจสอบความแตกต่างในภาพรวมจาก 5 ตัวแปร ได้แก่ เพศ การเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา เศรษฐานะของครอบครัว แหล่งทรัพยากรการศึกษาที่บ้าน และความมั่งคั่งของครอบครัว ระหว่างโมเดลประเมินคุณภาพการจัดการศึกษาในวิชาวิทยาศาสตร์ที่ไม่ได้พิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ (โมเดลที่ 1) และโมเดลการประเมินคุณภาพการจัดการศึกษาในวิชาวิทยาศาสตร์ที่พิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ โดยการตัดข้อสอบที่มีการทำหน้าที่ต่างกันของข้อสอบและการทำหน้าที่ต่างกันเป็นขั้นที่ต่างกัน (โมเดลที่ 2) พบว่า ค่าเฉลี่ยคะแนนการประเมินผลการรู้เรื่องวิทยาศาสตร์ระหว่างโมเดลที่ 1 และโมเดลที่ 2 มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ทั้ง 5 ตัวแปร ซึ่งผลการวิจัยดังกล่าวทำให้ทราบว่าคุณลักษณะของตัวแปรทั้ง 5 ตัวแปร ที่นำมาวิเคราะห์เมื่อนำมาทดสอบค่าเฉลี่ยระหว่างโมเดลทั้ง 2 ในแต่ละกลุ่มตัวแปรมีความแตกต่างกัน

ส่วนผลการเปรียบเทียบประสิทธิภาพการจัดการศึกษาระหว่าง โมเดล Undetected DIF-DSF-DTF (โมเดลที่ 1) กับ โมเดล Detected DIF-DSF-DTF (โมเดลที่ 2) โดยการตรวจสอบสัมประสิทธิ์ทำนายหรืออธิบายได้ (R^2) ของโมเดล Undetected DIF-DSF-DTF พบว่า กลุ่มตัวแปรระดับนักเรียนและกลุ่มตัวแปรบริบทระดับสถานศึกษาสามารถอธิบายสัดความแปรปรวนได้ถึง .49507 (49.507%) ส่วนโมเดล Detected DIF-DSF-DTF พบว่า กลุ่มตัวแปรระดับนักเรียนและกลุ่มตัวแปรบริบทระดับสถานศึกษาสามารถอธิบายสัดความแปรปรวนได้ถึง .49946 (49.946%) ดังนั้น ค่าเฉลี่ยรวมคะแนนการประเมินผลการรู้เรื่องวิทยาศาสตร์ที่มีการควบคุมตัวแปรระดับนักเรียนและสถานศึกษาหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน สามารถอธิบายสัดส่วนความ

แปรปรวนได้มากกว่าก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน สอดคล้องกับผลจากวิเคราะห์โมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันทั้งก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันในภาพรวมว่าทั้งสองโมเดลยังคงมีความตรงเชิงโครงสร้าง (construct validity) แต่เมื่อพิจารณาโมเดลก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า มีการปรับโมเดลในส่วนที่เป็นความสัมพันธ์ของคลาดเคลื่อนจนได้โมเดลที่เป็นไปตามค่าดัชนีตรวจสอบความสอดคล้อง จำนวน 2 คู่ แต่ว่าเมื่อพิจาณาโมเดลหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน พบว่า ไม่มีการปรับโมเดลในส่วนที่เป็นความสัมพันธ์ของคลาดเคลื่อน นอกจากนี้ผลการวิจัยดังกล่าวสอดคล้องกับผลการวิจัยทั้งของ Tuksino, Kanjanawasee, และ Pasiphol (2010) ซึ่งได้ประยุกต์ใช้โมเดลมูลค่าเพิ่ม (Value-added Model) ด้วยโมเดลเชิงเส้นตรงแบบลดหลั่น (HLM) ที่ได้จากการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบเพื่อการประเมินคุณภาพการจัดการศึกษาวิชาวิทยาศาสตร์ โดยการตรวจสอบผลการประเมินคุณภาพการจัดการศึกษาของสถานศึกษาจากโมเดลที่ต่างกัน 4 โมเดล พบว่า โมเดลประเมินคุณภาพการจัดการศึกษาวิชาวิทยาศาสตร์ที่ตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันออกและวิเคราะห์มูลค่าเพิ่มโดยการปรับแก้คะแนนด้วยตัวแปรระดับนักเรียนและสถานศึกษาที่มีประสิทธิ์ในการทำนายสูงสุด และ Sirikit, Pasiphol, และ Kanjanawasee (2012) ได้ประยุกต์ใช้โมเดลมูลค่าเพิ่ม ที่ได้จากการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและการวิเคราะห์การทำหน้าที่ต่างกันของตัวลวงเพื่อการประเมินคุณภาพการจัดการศึกษาวิชาคณิตศาสตร์ โดยการตรวจสอบผลการประเมินคุณภาพการจัดการศึกษาของสถานศึกษาจากโมเดลที่ต่างกัน 4 โมเดล พบว่า โมเดลการประเมินคุณภาพการจัดการศึกษาคณิตศาสตร์ที่ตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันโดยการวิเคราะห์ทั้งตัวถูกและตัวลวงออกและวิเคราะห์มูลค่าเพิ่มโดยการควบคุมตัวแปรระดับนักเรียนและระดับสถานศึกษาที่มีประสิทธิ์ในการทำนายสูงสุด

นอกจากการเปรียบเทียบประสิทธิภาพโมเดลมูลค่าเพิ่มพหุระดับแล้ว ยังมีประเด็นที่น่าสนใจอีกประเด็นคือ ตัวแปรที่นำมาศึกษาที่ส่งต่อการประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ทั้งในระดับนักเรียนและสถานศึกษา ซึ่งตัวแปรดังกล่าวยังส่งผลต่อคะแนนมูลค่าเพิ่มพหุระดับอีกด้วย รายละเอียดประเด็นตัวแปรทั้งในระดับนักเรียนและสถานศึกษาที่น่าสนใจมีดังต่อไปนี้

ประเด็นตัวแปรระดับนักเรียนที่ส่งผลต่อคะแนนมูลค่าเพิ่มวิทยาศาสตร์ ผลการวิเคราะห์มูลค่าเพิ่มพหุระดับจากการประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ ตัวแปรระดับนักเรียนที่นำมาศึกษาทั้งในโมเดลการประเมินคุณภาพการจัดการศึกษาวิชาวิทยาศาสตร์ที่ไม่ได้พิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ (โมเดลที่ 1) กับโมเดลการประเมินคุณภาพการจัดการศึกษาในวิชาวิทยาศาสตร์ที่พิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ โดยการตัดข้อสอบที่มีการทำหน้าที่ต่างกันของข้อสอบและการทำหน้าที่ต่างกันเป็นขั้นที่ต่างกัน (โมเดลที่ 2) จากตัวแปรทั้งหมด 6 ตัว พบว่า มีตัวแปรจำนวน 3 ตัวที่มีอิทธิพลต่อการประเมินคุณภาพการจัดการศึกษา

วิทยาศาสตร์ โดยตัวแปรที่มีอิทธิพลทางบวก คือ เศรษฐานะของครอบครัว และแหล่งทรัพยากรการศึกษาที่บ้าน ส่วนอิทธิพลทางลบ คือ ความมั่งคั่งของครอบครัว ตัวแปรดังกล่าวสอดคล้องกับงานวิจัยของ Caro (2009) ได้ดำเนินการวิจัย เศรษฐานะของครอบครัวของนักเรียนและโอกาสความไม่เท่าเทียมกันทางสังคม งานวิจัยดังกล่าวชี้ให้เห็นว่านักเรียนที่มาจากครอบครัวที่มีเศรษฐานะดีมักมีพัฒนาการทักษะการเรียนรู้รวดเร็วมากกว่านักเรียนที่มาจากครอบครัวที่มีเศรษฐานะไม่ดี ซึ่งผลที่ตามมาคือทำให้เกิดช่องว่างทางการเรียนรู้โดยพิจารณาจากผลสัมฤทธิ์ทางการศึกษาในโรงเรียน เช่นเดียวกับ Sun และคณะ (2012) นักเรียนที่มีเศรษฐานะทางครอบครัวที่ดี มีแรงกระตุ้นสูงและมีความเชื่อในประสิทธิภาพของตน (Self-efficacy) มักจะมีแนวโน้มจะมีผลสัมฤทธิ์ทางวิทยาศาสตร์จากผลการทดสอบ PISA สูงตามไปด้วย Johnson และคณะ (2007) พบว่าเศรษฐานะมีส่วนเกี่ยวข้องกับผลสัมฤทธิ์ทางการเรียนของนักเรียน โดยที่เศรษฐานะของผู้ปกครองมีอิทธิผลต่อศักยภาพการบริหารของสถานศึกษา Lee และ Wong (2004) พบว่า เชื้อชาติและเศรษฐานะส่งผลต่อผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ตามนโยบายการตรวจสอบได้ของรัฐ (Accountability policy) โดยอาศัยแนวคิดแบบการขับเคลื่อนตามสมรรถนะ(Performance-driven approach) และ Rumberger และ Palardy (2005) พบว่า ระดับเศรษฐานะเฉลี่ยของนักเรียนมีผลกระทบต่อพัฒนาการจากผลสัมฤทธิ์ทางการเรียน ส่วน Areepattamannil และ Kaur (2013) ได้ศึกษาองค์ประกอบที่ใช้ทำนายผลสัมฤทธิ์ทางวิทยาศาสตร์สำหรับนักเรียนที่ย้ายถิ่นมาจากต่างประเทศกับนักเรียนที่มีถิ่นอาศัยอยู่อาศัยในประเทศแคนาดา พบว่า ภาษาที่ใช้สื่อสารในครอบครัว ความมั่งคั่งของครอบครัว และเศรษฐานะทางครอบครัวของนักเรียนที่มีถิ่นอาศัยภายในประเทศล้วนเป็นองค์ประกอบที่มีผลต่อผลสัมฤทธิ์ทางวิทยาศาสตร์ ส่วนองค์ประกอบดังกล่าวไม่ได้มีความสัมพันธ์กับนักเรียนที่ย้ายถิ่นมาจากต่างประเทศ ส่วนทัศนคติต่อการเข้าศึกษา แรงจูงใจในวิชาวิทยาศาสตร์และเทคโนโลยีสารสนเทศและการสื่อสารเป็นปัจจัยที่มีต่อสัมฤทธิ์วิชาวิทยาศาสตร์ทั้งกลุ่มนักเรียนที่ย้ายถิ่นมาจากต่างประเทศกับนักเรียนที่มีถิ่นอาศัยอยู่อาศัยในประเทศ ในขณะที่การขาดแคลนครูมีความสัมพันธ์กับผลสัมฤทธิ์ทางการเรียนวิชาวิทยาศาสตร์กับกลุ่มนักเรียนที่ย้ายถิ่นมาจากต่างประเทศ ขนาดโรงเรียนมีความสัมพันธ์กับผลสัมฤทธิ์ทางการเรียนวิชาวิทยาศาสตร์นักเรียนที่มีถิ่นอาศัยอยู่อาศัยในประเทศ

ส่วนตัวแปรที่น่าสนใจอีกหนึ่งตัวแปรในระดับนักเรียน ก็คือ การใช้เวลาเรียนในการเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษาไม่มีผลต่อคะแนนมูลค่าเพิ่ม สอดคล้องการการวิจัยของ Tuksino, Kanjanawasee, และ Pasiphol (2010) พบว่า การใช้เวลาเรียนในการเรียนพิเศษวิทยาศาสตร์นอกสถานศึกษาไม่มีผลต่อการจัดคุณภาพการศึกษาวิชาวิทยาศาสตร์ อีกทั้งข้อสอบ PISA ยังเป็นข้อสอบที่เน้นสถานการณ์ที่นักเรียนต้องรู้จักการแก้ปัญหา ซึ่งอยู่นอกเหนือจากหลักสูตรการเรียนการสอนของการศึกษาในประเทศไทย และข้อสอบ PISA ยังเป็นข้อสอบที่กระตุ้นนักเรียน

ได้ฝึกการคิดวิเคราะห์ตามสถานการณ์ที่กำหนดให้ จำเป็นอย่างยิ่งที่นักเรียนจะต้องรู้จักการคิดวิเคราะห์และรู้จักการใช้เหตุผลในการตอบ นอกจากนี้ลักษณะข้อสอบมีทั้งแบบเลือกตอบและเขียนตอบ ในส่วนการเขียนตอบนั้นนักเรียนมีความจำเป็นอย่างยิ่งที่จะต้องมีทักษะในการเขียนและการสื่อความหมายให้ถูกต้อง ซึ่งเป็นทักษะที่นักเรียนทุกคนจะต้องฝึกปฏิบัติให้เกิดความชำนาญจากครูผู้สอนภายในโรงเรียนมาแล้ว ทำให้การกวดวิชาหรือการเรียนพิเศษไม่มีผลต่อการตอบข้อสอบ PISA เพราะ การกวดวิชาเป็นเพียงตัวเลือกของกลุ่มนักเรียนบางกลุ่มที่ต้องการเน้นในเนื้อหาในวิชาที่เรียนนอกเหนือจากห้องเรียนเพื่อสอบเท่านั้น และในส่วนของการจัดการเรียนการสอนในสถาบันกวดวิชาเองยังเป็นการสอนที่อิงเนื้อหาตามหลักสูตรของกระทรวงศึกษายังขาดการเชื่อมโยงไปยังเนื้อหาของข้อสอบ PISA ที่จัดสอบตามรอบที่กำหนดอีกด้วย

ประเด็นตัวแปรระดับสถานศึกษาที่ส่งผลต่อคะแนนมูลค่าเพิ่มวิทยาศาสตร์ ผลการวิเคราะห์มูลค่าเพิ่มพหุระดับจากการประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ ตัวแปรระดับสถานศึกษาที่นำมาศึกษาทั้งในโมเดลการประเมินคุณภาพการจัดการศึกษาวิชาวิทยาศาสตร์ที่ไม่ได้พิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ (โมเดลที่ 1) กับโมเดลการประเมินคุณภาพการจัดการศึกษาในวิชาวิทยาศาสตร์ที่พิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ โดยการตัดข้อสอบที่มีการทำหน้าที่ต่างกันของข้อสอบและการทำหน้าที่ต่างกันเป็นขั้นที่ต่างกัน (โมเดลที่ 2) จากตัวแปรทั้งหมด 6 ตัว พบว่า มีตัวแปรจำนวน 3 ตัวที่มีอิทธิพลต่อการประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ โดยที่ตัวแปรที่มี อิทธิพลทางบวก คือ ตัวแปรดัมมี่สถานศึกษาของรัฐ และทรัพยากรการเรียน ส่วนอิทธิพลทางลบ คือ ร้อยละงบประมาณของรัฐที่จัดสรรให้สถานศึกษา ตัวแปรดังกล่าวสอดคล้องกับงานวิจัยของ Lee และ Wong (2004) พบว่า เชื้อชาติและเศรษฐานะมีผลต่อร้อยละงบประมาณของรัฐที่จัดสรรให้สถานศึกษาแต่มีการจัดสรรเพิ่มขึ้นเพียงเล็กน้อยในเขตที่พบว่ามีโรงเรียนที่ยากจน เช่นเดียวกับงานวิจัยของ Ryan (2012) พบว่าความร่ำรวยของโรงเรียนมีความสัมพันธ์อย่างมีนัยสำคัญทางลบกับความสามารถของนักเรียน ส่วน Lubienski และ Lubienski (2006) ได้ทำการศึกษาผลสัมฤทธิ์ทางการเรียนของนักเรียนในโรงเรียนทางเลือก โรงเรียนของรัฐและเอกชนในประเทศสหรัฐอเมริกา พบว่า โรงเรียนเอกชนมีผลสัมฤทธิ์ทางการเรียนที่สูงกว่านักเรียนโรงเรียนของรัฐ และนักเรียนที่ด้อยโอกาสส่วนใหญ่มักจะเลือกเข้าศึกษาต่อในโรงเรียนของรัฐบาลมากกว่า และ Subedi (2003) ได้รายงานปัจจัยที่มีอิทธิผลต่อผลสัมฤทธิ์ทางการเรียนระดับมัธยมศึกษาในประเทศเนปาล พบว่า การใช้ทรัพยากรทางการศึกษามีต่อผลผลสัมฤทธิ์ทางการเรียนของนักเรียนภายในชั้นโดยขนาดชั้นเรียนมีอิทธิผลทางลบต่อระดับคะแนน

ส่วนตัวแปรที่น่าสนใจในระดับสถานศึกษา ก็คือ ระดับการขาดแคลนครู สัดส่วนนักเรียนต่อครู หลักสูตรการวัดและประเมินผล ไม่มีผลต่อคะแนนมูลค่าเพิ่มในการจัดการศึกษาของ

สถานศึกษา ปัจจุบันในประเทศไทยกำลังประสบปัญหาการขาดแคลนครูวิทยาศาสตร์ซึ่งสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (สพฐ.) ได้มีการเปิดเผยตัวเลขของจำนวนครูที่ขาดแคลน ในปีการศึกษา 2554 ในสาขาวิทยาศาสตร์ขาดแคลนครูจำนวน 6,815 อัตรา อยู่ในลำดับที่ 4 จาก 12 สาขาวิชา รองลงมาจาก ครูสาขาคณิตศาสตร์ สาขาภาษาต่างประเทศ และสาขาภาษาไทย ตามลำดับ (สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน, 2554) เป็นไปได้ว่าสถานศึกษามีการบริหารจัดการภายในองค์กรเพื่อชดเชยต่อสาขาที่ขาดแคลนโดยการใช้อัตรากำลังเสริมหรือที่เรียกว่าครูอัตราจ้าง หรือใช้ครูที่มีคุณวุฒิที่ใกล้เคียงเข้ามาสอนในเนื้อหาวิชาวิทยาศาสตร์ให้เหมาะสมกับจำนวนนักเรียนที่จะต้องเรียนในแต่ละปีการศึกษา นอกจากนี้การเพิ่มช่องทางอื่นๆ เช่น การสอนทางไกลแบบดาวเทียม การสอนผ่านโทรทัศน์หรือทางอินเตอร์เน็ต เช่น โทรทัศน์ครู (Thai Teachers TV) จากคณะกรรมการการอุดมศึกษา ก็เป็นอีกหนึ่งช่องทางเสริมเพื่อให้ครูหรือนักเรียนที่ต้องการความรู้เสริมเข้ามาศึกษาเนื้อหาวิชาที่ต้องการได้ แม้กระทั้งการจัดโครงการและกิจกรรมต่างๆ เพื่อส่งเสริมและสนับสนุนให้โรงเรียนมีการพัฒนาการเรียนการสอนกทำให้สถานศึกษามีการพัฒนาอย่างต่อเนื่อง เช่น โครงการโอลิมปิควิชาการ แข่งขันการตอบปัญหาวิทยาศาสตร์ทุกระดับ ซึ่งเป็นแรงกระตุ้นให้สถานศึกษาเกิดการพัฒนานักเรียนเพื่อเป็นตัวแทนให้เข้ามาแข่งขันทางด้านวิชาการให้เต็มศักยภาพ

5.2.2.4 ความสัมพันธ์ระหว่างโมเดลมูลค่าเพิ่มพหุระดับของ 2 โมเดล

ในส่วนนี้เป็นการวิเคราะห์ความสัมพันธ์ระหว่าง 4 ตัวแปร ได้แก่ คะแนนเฉลี่ยการประเมินผลการรู้เรื่องวิทยาศาสตร์ คะแนนมูลค่าเพิ่ม (VAS) การจัดอันดับคุณภาพคะแนนมูลค่าเพิ่ม (VAS ranking) และการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม (VAS rating) ของทั้งสองโมเดล พบว่า ทั้งสองโมเดลมีตัวแปรคะแนนเฉลี่ยการประเมินผลการรู้เรื่องวิทยาศาสตร์มีความสัมพันธ์กับคะแนนมูลค่าเพิ่ม (VAS) จัดการอันดับคุณภาพคะแนนมูลค่าเพิ่ม (VAS ranking) และการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม (VAS rating) อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนความสัมพันธ์ระหว่างคะแนนมูลค่าเพิ่ม (VAS) กับการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่ม (VAS ranking) มีความสัมพันธ์เกือบสมบูรณ์ในทิศทางตรงกันข้าม และความสัมพันธ์ระหว่างคะแนนมูลค่าเพิ่ม (VAS) กับการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม (VAS rating) มีความสัมพันธ์เกือบสมบูรณ์ในทิศทางเดียวกัน ส่วนความสัมพันธ์ระหว่างการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่ม (VAS ranking) และตัวการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม (VAS rating) มีความสัมพันธ์เกือบสมบูรณ์ในทิศทางตรงกันข้าม สอดคล้องกับงานวิจัยของ Hill และคณะ (2010) ยังได้ให้ระบุข้อดีเกี่ยวกับการตรวจสอบความสัมพันธ์ว่าเป็นหลักฐานที่มีความสำคัญชิ้นหนึ่งสำหรับเป็นการตรวจสอบความตรง

ของคะแนนมูลค่าเพิ่ม (value-added scores) ซึ่งความสัมพันธ์ที่สูงสามารถบอกได้ว่ามีความตรงเชิงลู่เข้า (convergent validity) และงานวิจัยของ เพ็ญภัคร พื้นผา (2554) ซึ่งได้มีการศึกษาความสัมพันธ์ระหว่างคะแนนมูลค่าเพิ่ม (VAS) กับอันดับประสิทธิผลของโรงเรียนที่จัดอันดับด้วยคะแนนมูลค่าเพิ่ม (VAS Ranks) พบว่า ความสัมพันธ์ระหว่างคะแนนมูลค่าเพิ่มกับอันดับประสิทธิผลของโรงเรียนที่จัดอันดับด้วยคะแนนมูลค่าเพิ่มมีความสัมพันธ์กันเกือบสมบูรณ์

นอกจากนี้ผู้วิจัยจึงได้วิเคราะห์ความสัมพันธ์ระหว่างคะแนนความสามารถ (theta scores) และคะแนนมูลค่าเพิ่ม (value-added scores) ตัวแปร 3 ตัวแปร ได้แก่ เศรษฐานะของครอบครัว แหล่งทรัพยากรการศึกษาที่บ้าน และความมั่งคั่งของครอบครัว ซึ่งเป็นตัวแปรต่อเนื่องระดับนักเรียนที่ถูกนำมาพิจารณาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ และการทำหน้าที่ต่างกันเป็นขั้น และเป็นตัวแปรควบคุมในโมเดลวิเคราะห์มูลค่าเพิ่มพหุระดับระหว่างโมเดล ที่ 1 กับโมเดลที่ 2 พบว่า ความสัมพันธ์ระหว่างคะแนนค่าความสามารถ (theta scores) กับ 3 ตัวแปร มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และความสัมพันธ์ระหว่างคะแนนมูลค่าเพิ่ม (value-added scores) กับทั้ง 3 ตัวแปร มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 เช่นเดียวกัน เมื่อพิจารณาความสัมพันธ์ระหว่างคะแนนความสามารถก่อนตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันกับ 3 ตัวแปร กับคะแนนมูลค่าเพิ่มหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันกับ 3 ตัวแปร พบว่า ค่าความสัมพันธ์ระหว่างคะแนนมูลค่าเพิ่มหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันลดลงทั้ง 3 ตัวแปร ซึ่งความสัมพันธ์ระหว่างคะแนนมูลค่าเพิ่มกับตัวแปรที่มีการควบคุมความคลาดเคลื่อนนั้นสามารถสะท้อนให้เห็นว่าแหล่งความคลาดเคลื่อนที่เกิดขึ้นนั้นมีทั้งในระดับผู้เรียนและสถานศึกษา ความไม่เท่าเทียมกันในสถานศึกษาอาจจะเกิดจากแหล่งความคลาดเคลื่อนที่มาจากต่างแหล่ง ทั้งนี้การนำแหล่งความคลาดเคลื่อนเหล่านั้นมาควบคุมทั้งในระดับผู้เรียนโดยการลดข้อสอบบางข้อที่ตรวจพบว่าเกิดความลำเอียงและสถานศึกษาโดยควบคุมตัวแปรก็จะได้คะแนนที่แท้จริง สอดคล้องกับ McCaffrey และคณะ (2003) ได้กล่าวว่า การลดความคลาดเคลื่อนที่เกิดขึ้นควรที่จะรวมตัวกันเป็นตัวแปรพยากรณ์ในโมเดลถดถอยมูลค่าเพิ่ม (value-added regression models) จะทำให้ได้คะแนนที่แท้จริงที่ได้มาจากการลดความคลาดเคลื่อน

5.2.2.5 เปรียบเทียบความสอดคล้องของประสิทธิผลของโรงเรียนการจัดอันดับคุณภาพและการจัดกลุ่มคุณภาพระหว่างโมเดลมูลค่าเพิ่มพหุระดับระหว่าง 2 โมเดล

ในส่วนผลการเปรียบเทียบความสอดคล้องของการจัดอันดับคุณภาพโรงเรียนจากคะแนนมูลค่าเพิ่มโดยสถิติทดสอบ Wilcoxon Signed Ranks พบว่า โมเดลที่ 1 และโมเดลที่ 2 มีการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่มไม่สอดคล้องกัน ผู้วิจัยให้ข้อสังเกตว่าจากการศึกษาคะแนน

มูลค่าในโมเดลที่ 2 เป็นแนวคิดที่นำความรู้ทางด้านวัดผลการศึกษาเข้ามาขจัดความลำเอียงของข้อสอบบางข้อออกไป เพื่อให้ได้คะแนนมูลค่าเพิ่มที่ใกล้เคียงกับความเป็นจริงมากที่สุด เพราะถ้าหากไม่ตัดข้อสอบบางข้อที่ตรวจพบการทำหน้าต่างกันออกไปอาจจะผลไปถึงการจัดลำดับ (ranking) หรือการจัดกลุ่มคุณภาพ (rating) ของโรงเรียนที่สามารถส่งผลไปถึงการตัดสินการประเมินคุณภาพสถานศึกษาผิดพลาดตามมา นอกจากนี้โมเดลที่ 2 ยังเป็นโมเดลที่นำแนวคิดการศึกษาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบเข้ามาช่วยลดคะแนนเฟ้อที่อาจจะส่งผลการต่อการประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ เมื่อนำผลการจัดอันดับมาเปรียบเทียบกันทำให้โมเดลที่ 1 และโมเดลที่ 2 มีผลการประเมินที่แตกต่างกัน

ส่วนผลการเปรียบเทียบความสอดคล้องของการจัดกลุ่มคุณภาพโรงเรียน 5 กลุ่ม ตามคะแนนมูลค่าเพิ่ม พบว่า โมเดลที่ 1 กับโมเดลที่ 2 มีการจัดกลุ่มคุณภาพโรงเรียนที่สอดคล้องกันอย่างมีนัยสำคัญทางสถิติ เบื้องต้นผู้วิจัยจึงได้ทดสอบความเท่ากันของค่าความแปรปรวน (levene's test) เพื่อตัดสินประสิทธิผลของโรงเรียน 2 กลุ่ม คือ กลุ่มโรงเรียนที่มีประสิทธิผล (effective school) และกลุ่มโรงเรียนที่ไม่มีประสิทธิผล (ineffective school) โดยที่เกณฑ์การพิจารณาโรงเรียนที่มีคะแนนมูลค่าเพิ่มตั้งแต่ 0 ขึ้นไป เป็นโรงเรียนที่มีประสิทธิผลส่วนโรงเรียนที่มีคะแนนมูลค่าเพิ่มน้อยกว่า 0 หรือติดลบ เป็นโรงเรียนที่ไม่มีประสิทธิผล จากผลการวิเคราะห์ทำให้สรุปได้ว่า โมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ที่ไม่ได้พิจารณาข้อสอบที่ทำหน้าที่ต่างกันข้อสอบและการทำหน้าที่ต่างกันของแบบสอบ (โมเดลที่ 1) และโมเดลการประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ที่พิจารณาข้อสอบที่ทำหน้าที่ต่างกันข้อสอบและการทำหน้าที่ต่างกันของแบบสอบ (โมเดลที่ 2) มีความแปรปรวนของคะแนนมูลค่าเพิ่มระหว่างกลุ่มโรงเรียนที่มีประสิทธิผลและโรงเรียนที่ไม่มีประสิทธิผลแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

ต่อผู้วิจัยจึงได้ทดสอบความสอดคล้องของการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม (VAS rating) ตามเกณฑ์การจัดกลุ่มคุณภาพการจัดการศึกษา 5 ระดับ คือ ระดับ A (ดีมาก) ระดับ B (ดี) ระดับ C (พอใช้) ระดับ D (ผ่าน) และระดับ E (ควรปรับปรุง) ซึ่งเป็นการพัฒนาเกณฑ์ในการเปรียบเทียบโมเดลขึ้นมาใหม่ โดยอาศัยหลักการและแนวคิดของนักวิชาการที่กล่าวไว้ว่า เกณฑ์หมายถึงระดับหรือมาตรฐานถือว่าเป็นความสำเร็จของการดำเนินงานที่ใช้ตัดสินคุณภาพหรือพฤติกรรมหรือคุณลักษณะที่ต้องการวัด อาจมีทั้งเกณฑ์ขั้นต่ำ ขั้นสูง ที่เป็นที่ยอมรับกันว่าเป็นมาตรฐาน ทั้งนี้เกณฑ์อาจมีการแปรเปลี่ยนไปไม่คงที่แน่นอนแล้วแต่สถานการณ์ตามลักษณะของกลุ่มและสภาพความจำเป็น (สิริกัญญา พวงสมบัติ, 2538; อุทุมพร จามรมาน, 2544; ศิริชัย กาญจนวาสี, 2554) ซึ่งระดับของเกณฑ์อาจจะมีได้หลากหลาย เช่น เกณฑ์ 4 ระดับ ได้แก่ ระดับดีเยี่ยม (outstanding) ระดับดีมาก (superior) ระดับดี (full satisfactory) และระดับพอใช้ (satisfactory) หรือ 5 ระดับ ได้แก่ ระดับความสามารถดีเลิศ ดี พอใช้ อ่อนตก ซึ่งในการวิจัยครั้งนี้ผู้วิจัยได้แบ่ง

เกณฑ์ตัดสินตามการจัดกลุ่มคุณภาพออกเป็น 5 ช่วง ได้แก่ กลุ่มโรงเรียนที่มีผลการประเมิน ระดับ A (ดีมาก) ระดับ B (ดี) ระดับ C (พอใช้) ระดับ D (ผ่าน) และระดับ E (ควรปรับปรุง) เพื่อระบุสภาพผลการประเมินสถาบันการศึกษาตามการจัดกลุ่มคุณภาพให้มีความชัดเจนยิ่งขึ้น ช่วยลดความคลาดเคลื่อนจากเกณฑ์การตัดสิน จากนั้นทดสอบการวัดความสอดคล้อง (Measure of Agreement) ของ Cohen's Kappa ซึ่งใช้ค่าสัมประสิทธิ์ของ Kappa เป็นตัวชี้วัดทางสถิติซึ่งประโยชน์ของ Cohen's kappa coefficient นั้นเป็นค่าสถิติที่ใช้ทดสอบความสอดคล้อง (consistency) กันของข้อมูล 2 กลุ่มและเป็นตัวบ่งชี้ระดับความเที่ยงจากผู้ให้คะแนนหลายคน (inter-rater reliability) (Cohen, 1960) ผลการวิจัย พบว่า โมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ตามการจัดกลุ่มคุณภาพระหว่าง โมเดลที่1 กับโมเดลที่ 2 มีการจัดกลุ่มคุณภาพโรงเรียนที่สอดคล้องกันระดับค่อนข้างสมบูรณ์ (almost perfect agreement) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 จากผลวิจัยดังกล่าวสอดคล้องกับจากการวิจัยของ Tuksino, Kanjanawasee, และPasiphol (2010) และSirikit, Pasiphol, และ Kanjanawasee (2012) ใช้เกณฑ์ในการจัดกลุ่มคุณภาพโดยพิจารณาจากค่าเปอร์เซ็นต์ไทล์ แบ่งเป็น 3 กลุ่ม คือ กลุ่มต่ำ (เปอร์เซ็นต์ไทล์ 1 – 24.99) กลุ่มปานกลาง (เปอร์เซ็นต์ไทล์ 25 – 74.99) และกลุ่มสูง (เปอร์เซ็นต์ไทล์ 76 – 100)

ทั้งนี้ผลการทดสอบความสอดสอดคล้องการจัดกลุ่มคุณภาพของสถานศึกษา พบว่า โมเดลที่ 1 สอดคล้องกับโมเดลที่ 2 โดยภาพรวมจำนวน 202 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 90.58 แยกพิจารณาตามตัวแปรสถานศึกษา ได้แก่ ขนาดสถานศึกษา ประเภทสถานศึกษา สังกัดของสถานศึกษา ตำแหน่งที่ตั้งของสถานศึกษา และระดับการขาดแคลนครูวิทยาศาสตร์ มีร้อยละความสอดคล้องตั้งแต่ร้อยละ 84.62 - 100 ดังนี้

ขนาดสถานศึกษาเมื่อแยกพิจารณาเป็น**ขนาดสถานศึกษาขนาดเล็ก** พบว่า จากจำนวนสถานศึกษา 169 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 152 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 89.94 **ขนาดสถานศึกษาขนาดกลาง** พบว่า จากจำนวนสถานศึกษา 51 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 47 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 92.16 และ**ขนาดสถานศึกษาขนาดใหญ่** พบว่า จากจำนวนสถานศึกษา 3 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 3 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 100

ประเภทสถานศึกษาเมื่อแยกพิจารณาเป็น**ประเภทสถานศึกษาของรัฐ** พบว่า จากจำนวนสถานศึกษา 197 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 178 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 90.36 และ**ประเภทสถานศึกษาเอกชน** พบว่า จากจำนวนสถานศึกษา 26 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 24 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 92.31

สังกัดของสถานศึกษาเมื่อแยกพิจารณาเป็น**โรงเรียนในสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานที่มาจากโรงเรียนขยายโอกาสเดิม (สพฐ.1)** พบว่า จากจำนวนสถานศึกษา 35 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 30 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 85.71 **โรงเรียนในสังกัดกรมสามัญศึกษาเดิม (สพฐ.2)** พบว่า จากจำนวนสถานศึกษา 91 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 82 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 90.11 **โรงเรียนในสังกัดสำนักบริหารงานคณะกรรมการส่งเสริมการศึกษาเอกชน (สช.)** พบว่า จากจำนวนสถานศึกษา 13 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 13 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 100.00 **โรงเรียนในสังกัดสำนักการศึกษากรุงเทพมหานคร (กทม.)** พบว่า จากจำนวนสถานศึกษา 20 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 19 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 95.00 **โรงเรียนในสังกัดสำนักประสานและพัฒนาการจัดการศึกษาท้องถิ่น (กศท.)** พบว่า จากจำนวนสถานศึกษา 17 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 16 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 94.12 **โรงเรียนในสังกัดสาธิตของมหาวิทยาลัย (สาธิต)** พบว่า จากจำนวนสถานศึกษา 16 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 15 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 93.75 **โรงเรียนอาชีวศึกษาของเอกชน (อศ.1)** พบว่า จากจำนวนสถานศึกษา 13 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 11 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 84.62 และ**วิทยาลัยอาชีวศึกษาของรัฐ (อศ.2)** พบว่า จากจำนวนสถานศึกษา 18 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 16 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 88.89

ตำแหน่งที่ตั้งของสถานศึกษาเมื่อแยกพิจารณาเป็น**สถานศึกษาที่ตั้งอยู่ในหมู่บ้าน** พบว่า จากจำนวนสถานศึกษา 47 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 40 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 85.11 **สถานศึกษาที่ตั้งอยู่ในเมืองเล็ก** พบว่า จากจำนวนสถานศึกษา 45 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 43 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 95.56 **สถานศึกษาที่ตั้งอยู่ในเมือง** พบว่า จากจำนวนสถานศึกษา 73 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 65 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 89.04 **สถานศึกษาที่ตั้งอยู่ในเมืองใหญ่** พบว่า จากจำนวนสถานศึกษา 38 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 34 โรงเรียน คิดเป็นร้อยละที่สอดคล้อง ร้อยละ 89.47 และ**สถานศึกษาที่ตั้งอยู่ในเมืองใหญ่มาก** พบว่า จากจำนวนสถานศึกษา 20 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 20 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 100.00

ระดับการขาดแคลนครูวิทยาศาสตร์ เมื่อแยกพิจารณาเป็น**ระดับไม่ขาดแคลนครูวิทยาศาสตร์** พบว่า จากจำนวนสถานศึกษา 86 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 77 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 89.53 **ระดับขาดแคลนครูวิทยาศาสตร์น้อย** พบว่า จากจำนวนสถานศึกษา 41 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 40 โรงเรียน คิดเป็นร้อยละที่

สอดคล้องร้อยละ 97.56 **ระดับขาดแคลนครูวิทยาศาสตร์เป็นบางส่วน** พบว่า จากจำนวนสถานศึกษา 81 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 72 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 88.89 และ**ระดับขาดแคลนครูวิทยาศาสตร์จำนวนมาก** พบว่า จากจำนวนสถานศึกษา 15 โรงเรียน ทั้งสองโมเดลสอดคล้องกัน 13 โรงเรียน คิดเป็นร้อยละที่สอดคล้องร้อยละ 86.67

นอกจากนี้ผู้วิจัยให้ข้อสังเกตในเรื่องการตัดสินใจเลือกใช้ผลการประเมินคุณภาพสถานศึกษาควรที่จะพิจารณาร่วมกันระหว่างการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่ม (VAS ranking) และการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม (VAS rating) เพราะการพิจารณาเฉพาะจัดอันดับคุณภาพคะแนนมูลค่าเพิ่มเพียงอย่างเดียวอาจไม่ได้รับสารสนเทศที่ไม่เพียงพอ เพราะเป็นเพียงการบอกตำแหน่งของสถานศึกษาที่มีลำดับอยู่ในอันดับต้นหรือท้าย ยังขาดรายละเอียดในเชิงคุณภาพที่มีความชัดเจนและสามารถให้รายละเอียดเป็นระดับมากขึ้น ซึ่งประโยชน์ของการนำการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่ม (VAS rating) เข้ามาร่วมตัดสินคุณภาพนั้นจะทำให้ผลการประเมินเพิ่มความน่าเชื่อถือและทำให้เกิดความชัดเจนมากยิ่งขึ้น

5.3 ข้อเสนอแนะ

ในการวิจัยเรื่องการวิเคราะห์เปรียบเทียบโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์โดยการประยุกต์ใช้โมเดลมูลค่าเพิ่มพหุระดับที่มีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ ผลที่ได้จากการวิจัยครั้งนี้สามารถก่อให้เกิดประโยชน์นานับประการต่อวงการการศึกษาของประเทศ ทั้งนี้ผู้วิจัยได้นำเสนอข้อเสนอแนะในการนำผลการวิจัยไปใช้ประโยชน์และข้อเสนอแนะสำหรับการวิจัยครั้งต่อไป ดังต่อไปนี้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

5.3.1 ข้อเสนอแนะในการนำผลการวิจัยไปใช้ประโยชน์

5.3.1.1 การใช้ประโยชน์ในระดับชาติ

การประเมินสถานศึกษาให้มีประสิทธิภาพควรนำแนวคิดความรู้ทางด้านวัดผลการศึกษาเข้ามาขจัดความลำเอียงของข้อสอบบางข้อออกไป และนำคะแนนมูลค่าเพิ่มที่สะท้อนคุณภาพที่แท้จริงซึ่งเกิดจากการควบคุมปัจจัยความคลาดเคลื่อนที่อาจจะส่งผลไปถึงการจัดลำดับ (ranking) หรือการจัดกลุ่มคุณภาพ (rating) ต่อการตัดสินการประเมินคุณภาพสถานศึกษาทั้งในระดับมหาวิทยาลัย วิทยาลัย และระดับโรงเรียนให้มีคุณภาพและมาตรฐานที่แท้จริง

5.3.1.2 การใช้ประโยชน์ในระดับเขตพื้นที่การศึกษา

โมเดลมูลค่าเพิ่มสามารถสะท้อนคุณภาพที่แท้จริงและประสิทธิภาพของสถานศึกษาที่สังกัดในแต่ละเขตพื้นที่การศึกษา เขตพื้นที่การศึกษาสามารถนำคะแนนมูลค่าเพิ่มมาวิเคราะห์หาสาเหตุพร้อมกับการวางนโยบายที่เกี่ยวข้องกับตัวแปรที่สะท้อนประสิทธิภาพของสถานศึกษา นอกจากนี้เขตพื้นที่สามารถให้ข้อมูลจากผลการประเมินว่ามีปัจจัยใดที่เป็นลักษณะเฉพาะซึ่งสามารถส่งผลต่อความคลาดเคลื่อนเพื่อเป็นการวางแผนสำหรับการออกแบบประเมิน

5.3.1.3 การใช้ประโยชน์ในระดับในระดับสถานศึกษา

การใช้คะแนนมูลค่าเพิ่มที่เกี่ยวข้องกับการประเมินประสิทธิภาพของสถานศึกษาจากตัวแปรที่เกี่ยวข้องคุณภาพการศึกษาของสถานศึกษาหรือนักเรียน สถานศึกษาสามารถนำตัวแปรที่เกี่ยวข้องกับผลการประเมินทั้งในระดับสถานศึกษาและนักเรียนมาวิเคราะห์ว่ามีปัจจัยใดที่ส่งผลต่อคุณภาพที่แท้จริงซึ่งสามารถสะท้อนได้ว่าตัวแปรดังกล่าวเป็นตัวแปรที่ส่งผลต่อประสิทธิภาพหรือเป็นตัวแปรที่ส่งผลทำให้เกิดความคลาดเคลื่อน

5.3.1.4 การใช้ประโยชน์ในระดับห้องเรียน

เนื่องจากคะแนนมูลค่าเพิ่มในการศึกษาครั้งนี้เป็นการศึกษาที่อาศัยหลักการทางด้านการวัดผลการศึกษาในส่วนการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ การนำแนวคิดในการขจัดความอยุติธรรมของข้อสอบที่เอนเอียงไปยังนักเรียนกลุ่มใดกลุ่มหนึ่งนั้นจำเป็นอย่างยิ่งต่อคะแนนที่แท้จริงของนักเรียนแต่ละคน โดยภาพรวมจะต้องมีการตรวจสอบการทำหน้าที่ต่างกันของแบบสอบร่วมกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทุกครั้งก่อนที่ครูจะตัดข้อสอบที่ตรวจพบว่าข้อสอบบางข้อทำหน้าที่ต่างกันออกไปซึ่งจะช่วยเพิ่มความตรงของข้อสอบให้เหมาะสมกับกลุ่มนักเรียนมากขึ้น

5.3.2 ข้อเสนอแนะสำหรับการวิจัยครั้งต่อไป

ผลจากการวิจัยครั้งนี้ ผู้วิจัยได้นำเสนอองค์ความรู้ที่ได้จากการศึกษาการวิเคราะห์โมเดลมูลค่าเพิ่มพหุระดับโดยการประยุกต์ใช้การทำหน้าที่ต่างกันของข้อสอบและการทำหน้าที่ต่างกันของแบบสอบตามบริบทการศึกษาของประเทศไทย ซึ่งส่วนหนึ่งของผลการวิจัยจำเป็นที่จะต้องมีการศึกษาในรายละเอียดต่างๆ เพิ่มเติมในหลายประเด็น ทั้งนี้เพื่อเป็นการพัฒนาองค์ความรู้ในด้านการวัดผลให้ลุ่มลึกและเป็นการขยายองค์ความรู้ให้พัฒนากว้างขวางมากยิ่งขึ้น ผู้วิจัยขอเสนอแนะประเด็นที่ควรดำเนินการวิจัยต่อไป ดังนี้

1) การพิจารณาประสิทธิภาพของครูและโรงเรียนควรที่จะมีการศึกษากับข้อมูลระยะยาว (longitudinal data) และเพิ่มระดับการวิเคราะห์ให้มีความซับซ้อน เป็น 4 ระดับ ดังนี้ ระดับที่ 1 ระดับนักเรียน (student level) ระดับที่ 2 ระดับครู (teacher level) ระดับที่ 3 ระดับสถานศึกษา (school level) และระดับที่ 4 ระดับเขตพื้นที่การศึกษา (education service area level)

2) โมเดลการวิเคราะห์มูลค่าเพิ่มพหุระดับควรมีการศึกษาประสิทธิภาพของสถานศึกษาให้ครอบคลุมและหลากหลายวิชาเพิ่มขึ้น เช่น การรู้เรื่องคณิตศาสตร์ การรู้เรื่องการอ่าน เพื่อนำผลการประเมินประสิทธิภาพของสถานศึกษามาปรับปรุงและพัฒนากระบวนการจัดการเรียนการสอนให้สอดคล้องกับบริบทของสถานศึกษา

3) เนื่องจากการวิจัยครั้งนี้เป็นการใช้ข้อมูลทุติยภูมิ และได้มีการคัดเลือกตัวแปรที่นำมาศึกษาเพื่อนำมาใช้ประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ ดังนั้นเพื่อทำให้เกิดประสิทธิภาพของสถานศึกษาอย่างแท้จริงควรที่จะมีการศึกษาปัจจัยเชิงสาเหตุเพิ่มเติมที่เกี่ยวข้องกับผลสัมฤทธิ์ทางการเรียนของนักเรียน เพื่อให้ได้ข้อมูลสารสนเทศที่เป็นประโยชน์ต่อการประเมินคุณภาพการจัดการศึกษาและการวิจัยเพิ่มขึ้น

4) การเปรียบเทียบโมเดลประสิทธิภาพการจัดการศึกษาควรที่จะเพิ่มเติมให้มีการศึกษาจากผลการวิเคราะห์ทางสถิติโดยการทดสอบความสอดคล้องของโมเดล (measures of fit: MOF) ด้วยวิธีการวัดค่าความสอดคล้องแบบการวัดเชิงตรวจสอบ (predictive measures) เพื่อเป็นการวัดเชิงตรวจสอบว่าโมเดลจะมีความคงเส้นคงวาเพียงใด จากค่าสถิติ AIC (Akaike information criterion) และ BIC (Bayesian information criterion)

5) การทดสอบความไม่แปรเปลี่ยนของโมเดล (measurement invariance) ในระดับแบบสอบตามองค์ประกอบที่ศึกษากับตัวแปรที่มีศึกษา เพื่อเป็นการศึกษาในเชิงลึกของ การวิเคราะห์โมเดลทั้งก่อนและหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบ

รายการอ้างอิง

ภาษาไทย

กาญจนา วัธนสุนทร. (2537). *การพัฒนาเกณฑ์การตัดสินข้อสอบลำเอียงทางเพศ*. วิทยานิพนธ์ปริญญาครุศาสตรดุษฎีบัณฑิต, จุฬาลงกรณ์มหาวิทยาลัย, ครุศาสตร์.

จตุภูมิ เขตจตุจัส. (2552). *การพัฒนโมเดลมูลค่าเพิ่มของผลสัมฤทธิ์ทางวิชาการและแบบตรวจสอบรายการประเมินตนเองเพื่อเพิ่มมูลค่ากระบวนการจัดการศึกษา*. วิทยานิพนธ์ปริญญาครุศาสตร์ดุษฎีบัณฑิต, จุฬาลงกรณ์มหาวิทยาลัย.

จารุวรรณ นาคคูบัว. (2552). *การศึกษาปัจจัยด้านเชาว์ปัญญา แรงจูงใจใฝ่สัมฤทธิ์ และรูปแบบการเรียนรู้ที่ส่งผลต่อผลการเรียนรู้วิชาคณิตศาสตร์และวิชาการไทย ของนักเรียนชั้นมัธยมปีที่ 3 สังกัดสำนักงานเขตพื้นที่การศึกษานครปฐม เขต 1.* วิทยานิพนธ์ปริญญามหาบัณฑิต, มหาวิทยาลัยศรีนครินทรวิโรฒ, คณะศึกษาศาสตร์.

ชลี ภัทรพิชญธรรม. (2553). *การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทางด้วยโมเดลเชิงเส้นตรงทั่วไประดับลดหลั่น: การประมาณค่าพารามิเตอร์.* วิทยานิพนธ์ปริญญาครุศาสตรดุษฎีบัณฑิต, จุฬาลงกรณ์มหาวิทยาลัย.

ทวีศิลป์ สารแสน. (2543). *ความสัมพันธ์ระหว่างองค์ประกอบของสภาพแวดล้อมทางการเรียนในห้องด้านครูผู้สอนกับความพึงพอใจของนักเรียนมัธยมศึกษา*. วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต, มหาวิทยาลัยศรีนครินทรวิโรฒ, คณะศึกษาศาสตร์.

ธเกียรติกมล ทองงอก, โชติกา ภาษีผล, และ ศิริชัย กาญจนวาสี. (2556). *ประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในวิธีถดถอยโลจิสติกโดยใช้เกณฑ์ขนาดอิทธิพล 2 วิธี สำหรับข้อสอบที่มีรูปแบบการตรวจให้คะแนนแบบทวิภาค: ข้อมูลจำลองและข้อมูลเชิงประจักษ์.* วารสารวิจัย มสด สาขามนุษยศาสตร์และสังคมศาสตร์, 9(2), 31-50.

นงลักษณ์ วิรัชชัย. (2542). *พรมแดนความรู้ด้านวิจัยและสถิติ รวมบทความวิชาการของ ดร.นงลักษณ์ วิรัชชัย*. วิทยาลัยการบริการรัฐกิจ มหาวิทยาลัยบูรพา. ชลบุรี: บริษัท เอ็ม เอ็นคอมพิวออฟเซท จำกัด.

บุญเรียง ขจรศิลป์. (2555). *การใช้โปรแกรม HLM ในการวิเคราะห์พหุระดับและการตีความผลที่ได้.* กรุงเทพฯ: โรงพิมพ์ พีเอส.พริ้นท์.

รัชนก บุญปู่. (2547). *ปัจจัยที่มีอิทธิผลต่อผลการเรียนรู้ตามมาตรฐานหลักสูตรการศึกษาขั้นพื้นฐานในโรงเรียนนำร่องหลักสูตรสถานศึกษา: การวิเคราะห์พหุระดับ*. วิทยานิพนธ์ปริญญาครุศาสตรดุษฎีบัณฑิต, จุฬาลงกรณ์มหาวิทยาลัย, คณะครุศาสตร์.

วลีมาศ แซงอึ้ง. (2543). *การเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิปเทสท์ปรับใหม่ วิธีซิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก.* วิทยานิพนธ์ปริญญาครุศาสตรดุษฎีบัณฑิต, จุฬาลงกรณ์มหาวิทยาลัย, ครุศาสตร์.

วัยวุฒ์ อยู่ในศิล. (2553). *การประยุกต์ใช้การวิเคราะห์สายโซ่วิธีการ-เป้าหมายในการพัฒนาคุณภาพการเรียนการสอนที่เน้นผู้เรียนเป็นศูนย์กลาง: การพัฒนาโมเดลการวัดและการวิเคราะห์มูลค่าเพิ่ม.* วิทยานิพนธ์ปริญญาครุศาสตรดุษฎีบัณฑิต, จุฬาลงกรณ์มหาวิทยาลัย, ครุศาสตร์.

ศักดิ์ชัย จันทะแสง. (2550). *การศึกษาปัจจัยด้านสติปัญญาและด้านที่ไม่ใช่สติปัญญาที่ส่งผลต่อผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์และวิชาวิทยาศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 3.* วิทยานิพนธ์ปริญญามหาบัณฑิต, มหาวิทยาลัยศรีนครินทรวิโรฒ, คณะศึกษาศาสตร์.

ศิริชัย กาญจนาวาสี. (2544). *ทฤษฎีการทดสอบแบบดั้งเดิม (*Classical Test Theory) (*พิมพ์ครั้งที่ 4)*. กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

ศิริชัย กาญจนาวาสี. (2550ก). *ทฤษฎีการทดสอบแนวใหม่ (พิมพ์ครั้งที่ 3).* กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

ศิริชัย กาญจนาวาสี. (2550ข). *การวิเคราะห์พหุระดับ (พิมพ์ครั้งที่ 4)*. กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

ศิริชัย กาญจนาวาสี. (2554). *ทฤษฎีการประเมิน (พิมพ์ครั้งที่ 7).* กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์กรมหาชน). (2553ก). *คู่มือการจัดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ช่วงชั้นที่ 4 (ม.6) ปีการศึกษา 2553*. เข้าถึง 25 พฤษภาคม 2554, มาจาก สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์กรมหาชน) http://www.niets.or.th/

สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์กรมหาชน). (2553ข). *คู่มือการจัดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) ช่วงชั้นที่ 4 (ม.6) ปีการศึกษา 2553.* เข้าถึง 25 พฤษภาคม 2554, จาก http://www.niets.or.th/

สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน). (2555). *รายงานประจำปี 2555.* เข้าถึง 6 พฤศจิกายน 2557, จาก www.niets.or.th/index.php/files/download/392/

สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี. (2551). *ตัวอย่างการประเมินผลวิทยาศาสตร์นานาชาติ* PISA *และ* TIMSS. กรุงเทพมหานคร: ห้างหุ้นส่วนจำกัด อรุณการพิมพ์.

สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี. (2553ก). *ผลการประเมิน* PISA *2009 การอ่าน คณิตศาสตร์ และวิทยาศาสตร์ บทสรุปเพื่อการบริหาร.* กรุงเทพมหานคร: สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี. .

สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี. (2553ข). *ผลการประเมิน* PISA *2009 การอ่าน คณิตศาสตร์ และวิทยาศาสตร์ บทสรุปเพื่อการบริหาร.* กรุงเทพมหานคร: สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี.

สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี. (2554). *ปัจจัยที่ทำให้ระบบโรงเรียนประสบความสำเร็จ ข้อมูลพื้นฐานจาก โครงการ* PISA *2009*. กรุงเทพมหานคร: สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี.

สัมพันธ์ พันธุ์พฤกษ์. (2557). *สอบ O-Net ปี 59 ใช้แนวข้อสอบ PISA Liked เน้นคิดวิเคราะห์*. เข้าถึง 12 พฤศจิกายน, 2557, จาก http://www.manager.co.th/OOL/ViewNews.aspx?NewsID=9570000102640/

สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน. (2554). *วิกฤตหนัก! ไทยขาดแคลนครูทุกวิชาเกือบ 6 หมื่นอัตรา.* เข้าถึง 28 ตุลาคม 2557, จาก http://education.kapook.com/view46762.html/

สิริกัญญา พวงสมบัติ. (2538). *การพัฒนาเกณฑ์การประเมินโครงการสุขภาพด้านการจัดบริการสุขภาพในโรงเรียนมัธยมศึกษา*. วิทยานิพนธ์ปริญญามหาบัณฑิต, จุฬาลงกรณ์มหาวิทยาลัย , คณะครุศาสตร์.

สุชีรา มะหิเมือง. (2547). *ปัจจัยที่ส่งผลต่อผลสัมฤทธิ์และพัฒนาการทางวิชาการ: การวิเคราะห์มูลค่าเพิ่ม.* วิทยานิพนธ์ปริญญาครุศาสตรดุษฎีบัณฑิต, จุฬาลงกรณ์มหาวิทยาลัย, คณะครุศาสตร์.

สุบิน ยุระรัช. (2547). *การสร้างโมเดลสามระดับของข้อมูลตัวแปรที่ส่งผลต่อคุณภาพนักเรียนประถมศึกษาในกรุงเทพมหานครโดยใช้โปรแกรมเอ็มแอลวิน*. วิทยานิพนธ์ปริญญาครุศาสตรดุษฎีบัณฑิต, จุฬาลงกรณ์มหาวิทยาลัย, คณะครุศาสตร์.

สุพักตร์ พิบูลย์. (2557). *การทดสอบระดับชาติ O-NET : หลักการสำคัญ คืออย่างไร.* เข้าถึง 25 กุมภาพันธ์, 2014

อภิเชษฐ์ ฉิมพลีสวรรค์. (2552). *การวิเคราะห์ปัจจัยเชิงสาเหตุของคุณภาพการจัดการศึกษาของโรงเรียนขนาดเล็ก สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน.* วิทยานิพนธ์ปริญญาครุศาสตรดุษฎีบัณฑิต, จุฬาลงกรณ์มหาวิทยาลัย, คณะครุศาสตร์.

อุทุมพร จามรมาน. (2544). *รายงานวิจัยของโครงการวิจัยและพัฒนาเรื่องดัชนีเกณฑ์และวิธีการประเมินคุณภาพการศึกษาภายในและภายนอกสำหรับสถาบันอุดมศึกษาไทย.* คณะครุศาสตร์: จุฬาลงกรณ์มหาวิทยาลัย.

ภาษาอังกฤษ

Adam, R. J. (2003). Response to " Cautions on OECD's Recent Educational Survey (PISA)" (Vol. 29, pp. 379-389). Oxford Review of Education.

Adam, R. J., Wu, M. L., & Carstensen, C. H. (2007). Application of Multivariate Rash models in International Large-Scale Educational Assesments. *Statistics for social and Behavioral Sciences*. New York: Springer.

American Statistical Association (2014). Using Value-added Models for Educational Assessment.

Amrein-Beardsley, A. (2008). Methodological Concerns About the Education Value-Added Assessment System. *Educational Researcher,* 37(2), 65-75. doi: 10.3102/0013189x08316420

Areepattamannil, S., & Kaur, B. (2013). FACTORS PREDICTING SCIENCE ACHEVEMENT OF IMMIGRANT AND NON- IMMIGRANT STUDENTS: A MULTILEVEL ANALYSIS. *International Journal of Science and Mathematics Education,* 11, 1183-1207.

Ayala, R. J. d. (2008). *The Theory and practice of item response theory*. New York: The Guilford press.

Bao, H., Dayton, C. M. & Hendrickson, A. B. (2009). Detecting Differential Item Functioning Amplification and Cancellation in a Reading Test. *Practical Assessment, Research & Evaluation,* 14(9), 1-27.

Barnes, B. J., & Wells, C. S. (2009). Differential Item Functional Analysis by Gender and Race of the National Doctoral Program Survey. *International Journal of Doctoral Studies,* 4, 77-96.

Beardsley, A. A. (2014). Methodological Concerns About the Education Value-Added Assessment System. *Educational Researcher,* 37(2), 65-75.

Camilli, G., & Congdon, P. (1999). Application of a Method of Estimating DIF for Polytomous Test Items. *Journal of Educational and Behavioral Statistics,* 24(4), 323-341. doi: 10.2307/1165366

Caro, D. H. (2009). *Family Socioeconomic Status and Inequality of Opportunity.* (Mater in Interdisciplinary Studies), Berlin.

Carvajal, J., & Skorupski, W. P. (2010). The Effects of Small Sample Size on Identifying Polytomous DIF Using the Liu-Agresti Estimator of the Cumulative Common Odds Ratio. *Educational and Psychological Measurement,* 70(6), 914-925. doi: 10.1177/0013164410379325

Chiu, M. M., & Xihua, Z. (2008). Family and motivation effects on mathematics achievement: Analyses of students in 41 countries. *Learning and Instruction,* 18(4), 321-336. doi: http://dx.doi.org/10.1016/j.learninstruc.2007.06.003

Cho, Y. I., Martin, M. J., Conger, R. D., & Widaman, K. F. (2010). Differential Item Functioning on Antisocial Behavior Scale Items for Adolescents and Young Adults from Single-Parent and Two-Parent Families. *Journal of Psychopathology and Behavioral Assessment,* 32(2), 157-168. doi: 10.1007/s10862-009-9145-1

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement,* 20(1), 37-46. doi: 10.1177/001316446002000104

Demar, c. (2010). *Item Response Theory (Understanding Statistics Measurement).* New York: Oxford University Press, Inc.

Downes, D., & Vindurampulle, O. (2007). Value-added measures for school improvement (Vol. 13). Melbourne: Education Policy and Research Division. Retrieved from https://www.eduweb.vic.gov.au/edulibrary/public/publ/research/publ/value-added-measures-report.pdf.

Drasgow, F., & Schmitt, N. (2002). *Measuring and Analyzing Behavior in Organizations*. New York: Jossey-Bass Inc.

Drury, D., & Doran, H. (2003). The Value of Value-Added Analysis. *Policy Research Brief,* 3(1), 1-4.

Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. New jersey: Lawrence Erlbaum associated publishers.

Fidalgo, A. M., & Bartram, D. (2010). A Comparison Between Some Generalized Mantel-Haenszel Statistics for Detecting DIF in Data Simulated Under the Graded Response Model. *Applied Psychological Measurement,* 34(8), 600-606. doi: 10.1177/0146621610378405

Finch, W. H., & B.F., F. (2007). Detection of Crossing Differential Item Functioning : Comparison of Four Methods. *Educational and Psychological Measurement,* 67(4), 565-582.

Fleischman, H. L., Hopstock, P. J. , Pelczar, M. P.,& Shelley B. E. (2010). *Highlights From PISA* 2009: *Performance of U.S.* 15*-Year-Old Students in Reading, Mathematics, and Science Literacy in an International Context* National Center of Education Statistics.

Gallagher, A. M., & Kaufman, J. C. (2005). *Gender Differences in Mathematics*. United Kingdom: Cambridge University Press.

Gattamorta, K. A., & Penfield, R. D. (2012). A Comparison of Adjacent Categories and Cumulative Differential Step Functioning Effect Estimators. *Applied Measurement in Education,* 25(2), 142-161. doi: 10.1080/08957347.2012.660387

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference.* 11.0 *update (*4*th ed.)*. Boston: Allyn & Bacon.

Geske, A., & Ozola, A. (2010). *Differential Item Functioning in the Aspect of Gender Differences in Reading Literacy*. Paper presented at the IEA International Research Conference.

Gideon, J. M. (1982). Contingency Table Models for Assessing Item Bias. *Journal of Educational Statistics,* 7(2), 105-118. doi: 10.2307/1164960

Gong, J. (2012). *Detection of Differential Test Functioning (DTF) and Differential Item Functioning (DIF) in MCCQE Part II Using Logistic Models.* Retrieved from http://mcc.ca/wp-content/uploads/technical-reports-gong-2012.pdf

Hair et al. (1998). *Multivariate Data Analysis*. New Jersey: Prentice-Hall PTR.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1985). *Item Response Theory: Principle and applications*. Boston: Kluwer Nijhoff Publishing.

Heck, R. H. (2000). Examining the Impact of School Quality on School Outcomes and Improvement: A Value-Added Approach. *Educational Administration Quarterly,* 36(4), 513-552. doi: 10.1177/00131610021969092

Holland, P. W., & Wainer, H. (1993). *Differential Item Functioning*. Hillsdale, Newjersy: Lawrence Erlbaum Associates.

Holland, P. W., & Thayer, D. T. (2000). Univariate and Bivariate Loglinear Models for Discrete Test Score Distributions. *Journal of Educational and Behavioral Statistics,* 25(2), 133-183. doi: 10.3102/10769986025002133

Holmes Finch, W., & French, B. F. (2007). Detection of Crossing Differential Item Functioning: A Comparison of Four Methods. *Educational and Psychological Measurement,* 67(4), 565-582. doi: 10.1177/0013164406296975

Hoxby, C. M. (2003). *The Economics of School Choice* (C. M. Hoxby Ed.). United states of America: University of Chicago Press.

Hulin, C. L., Drasgow, F. & Parson, C. K. (1983). *Item response theory : Applications of psychological measurement*. IL: Dow Jones Irwin.

Keeves, J. P., Hungi, N., & Afrassa, T. (2005). Measuring value added effects across schools: Should schools be compared in performance? *Studies in Educational Evaluation,* 31(2–3), 247-266. doi: http://dx.doi.org/10.1016/j.stueduc.2005.05.012

Kim, S., Cohen A. S., Alagoz, C. & Kim S. (2007). DIF Detection and Effect Size Measures for Polytomously Scored Items. *Journal of Educational Measurement,* 44(2), 93-116.

Kunnan, A. J. (2000). *Fairness and validation in language assessment: Selected papers from the* 19*th Language Testing Research Colloquium, Orlando, Florida*. N.Y: Press Syndicate of the University of Cambridge.

Lambert, D. L., D. . (2000). *Understanding assessment, purpose, perceptions, Practice*. Great Britain: TJ International Ltd. Padstow Cornwall.

Le, L. T. (2009). Investigating Gender Differential Item Functioning Across Countries and Test Languages for PISA Science Items. *International Journal of Testing,* 9(2), 122-133. doi: 10.1080/15305050902880769

Li, H.-H., & Stout, W. (1996). A new procedure for detection of crossing DIF. *Psychometrika,* 61(4), 647-677. doi: 10.1007/BF02294041

Lissitz, R. W. (2006). *Longitudinal and Value Added Models of Student Performance*. Minnesota: JAM Press.

Liu, O. L., & Wilson, M. (2009). Gender Differences in Large-Scale Math Assessments: PISA Trend 2000 and 2003. *Applied Measurement in Education,* 22(2), 164-184. doi: 10.1080/08957340902754635

McEwen, N. (1995). Accountability in Education in Canada. *Canadian Journal of Education,* 20(1), 1-107.

McNamara, T. & Roever, C. (2004). *Language Testing: The Social Dimension*. Malden: Blackwell Publishing.

Miller, T., Chahine, S. & Childs, R. A. (2010). Detecting Differential Item Functioning and Differential Step Functioning Due to Differences that Should Matter. *Practical Assessment Research & Evaluation,* 15, 1-13.

Narayanan, P., & Swaminathan, H. (1994). Performance of the Mantel-Haenszel and Simultaneous Item Bias Procedures for Detecting Differential Item Functioning. *Applied Psychological Measurement,* 18(4), 315-328. doi: 10.1177/014662169401800403

Noell, G. H., & Burns, J. L. (2006). Value-Added Assessment of Teacher Preparation: An Illustration of Emerging Technology. *Journal of Teacher Education,* 57(1), 37-50. doi: 10.1177/0022487105284466

OECD. (2008). *Measuring Improvements in Learning Outcomes: Best Practices to Assess the Value-Added of Schools*. Paris: OECD.

OECD. (2012). PISA 2009 Technical Report: OECD Publishing.

Oliveri, M. E., Olson, B. F., Ercikan, K., & Zumbo, B. D. (2012). Methodologies for Investigating Item- and Test-Level Measurement Equivalence in International Large-Scale Assessments. *International Journal of Testing,* 12(3), 203-223. doi: 10.1080/15305058.2011.617475

Penfield, R. D. (2005). DIFAS: Differential Item Functioning Analysis System. Computer Program Exchange. *Applied Psychological Measurement,* 29(2), 150-151.

Penfield, R. D. (2007a). An Approach for Categorizing DIF in Polytomous Items. *Applied Measurement in Education,* 20(3), 335-355. doi: 10.1080/08957340701431435

Penfield, R. D. (2007b). Assessing Differential Step Functioning in Polytomous Items Using a Common odds Ratio Estimator. *Journal of Educational Measurement,* 44(3), 187-210.

Penfield, R. D. (2008). Three Classes of Nonparametric Differential Step Functioning Effect Estimators. *Applied Psychological Measurement,* 32(6), 480-501. doi: 10.1177/0146621607305399

Penfield, R. D. (2010a). Distinguishing between Net and Global DIF in Polytomous. *Journal of Educational Measurement,* 47(2), 129-149.

Penfield, R. D. (2010b). Explaining Crossing DIF in Polytomous Items Using Differential Step Functioning Effects. *Applied Psychological Measurement,* 34(8), 563-579. doi: 10.1177/0146621610377083

Penfield, R. D. (2010c). How Are the Form and Magnitude of DIF Effects in Multiple-Choice Items Determined by Distractor-Level Invariance Effects? *Educational and Psychological Measurement,* 71(1), 54-67. doi: 10.1177/0013164410387340

Penfield, R. D. (2010d). Modeling DIF Effects Using Distractor-Level Invariance Effects: Implications for Understanding the Causes of DIF. *Applied Psychological Measurement,* 34(3), 151-165. doi: 10.1177/0146621609359284

Penfield, R. D., & Algina, J. (2003). Applying the Liu-Agresti Estimator of the Cumulative Common Odds Ratio to DIF Detection in Polytomous Items. *Journal of Educational Measurement,* 40(4), 353-370. doi: 10.2307/1435386

Penfield, R. D., & Algina, J. (2006). A Generalized DIF Effect Variance Estimator for Measuring Unsigned Differential Test Functioning in Mixed Format Tests. *Journal of Educational Measurement,* 43(4), 295-312. doi: 10.1111/j.1745-3984.2006.00018.x

Penfield, R. D., Alvarez, K., & Lee, O. (2008). Using a Taxonomy of Differential Step Functioning to Improve the Interpretation of DIF in Polytomous Items: An Illustration. *Applied Measurement in Education,* 22(1), 61-78. doi: 10.1080/08957340802558367

Penfield, R. D., Gattamorta, K. & Childs, R. A. (2009). An NCME Instructional Module on Using Differential Step Functioning ot Refine the Analysis of DIF in Polytomous Items. *Educational Measurement: Issues and Practice,* 28(1), 38-49.

Penfield, R. D., & GÜler, N. (2009). A Comparison of the Logistic regression and Contigency Talbe Methods for Simutaneous Detection of Uniform and Nonuniform DIF. *Journal of Educational Measurement,* 46(3), 314-329.

Penfield, R. D., Myers, N. D., & Wolfe, E. W. (2008). Methods for Assessing Item, Step, and Threshold Invariance in Polytomous Items Following the Partial Credit Model. *Educational and Psychological Measurement,* 68(5), 717-733. doi: 10.1177/0013164407312602

Perrone, M. (2006). Differential Item Functioning and Item Bias: Critical Considerations in Test Fairness. *TESOL & Applied Linguistics,* 6(2), 1-3.

Ready, D. D. (2013). Associations Between Student Achievement and Student Learning: Implications for Value-Added School Accountability Models. *Educational Policy,* 27(1), 92-120. doi: 10.1177/0895904811429289

Rogers, H. J. (2005). Differential Item Functioning *Encyclopedia of Statistics in Behavioral Science*: John Wiley & Sons, Ltd.

Sanders, W., & Horn, S. (1998). Research Findings from the Tennessee Value-Added Assessment System (TVAAS) Database: Implications for Educational Evaluation and Research. *Journal of Personnel Evaluation in Education,* 12(3), 247-256. doi: 10.1023/A:1008067210518

Shealy, R., & Stout, W. F. (1993). *Differential item functioning: Theory and practice*. New Jersey: Lawrence Erlbaum.

Sirikit, R., Pasiphol, C., & Kanjanawasee, S. (2012). A Comparative Analysis of the Model of A Quality Assessment of Mathematics Subjects: An Application of Differential Item Functioning and Differential Distractor Functioning. *Scholar,* 4(2), 73-77.

Steedel, J. T. (2010). Improving the Reliability and Interpretability of Value-Added Scores for Post-Secondary Institutional Assessment Program. *Presented at the* 2010 *annual Meeting of the American Educational Research Associaiton*, 1-18.

Stout, W., Li, H.-H., Nandakumar, R., & Bolt, D. (1997). MULTISIB: A Procedure to Investigate DIF When a Test is Intentionally Two-Dimensional. *Applied Psychological Measurement,* 21(3), 195-213. doi: 10.1177/01466216970213001

Swaminathan, H., & Rogers, H. J. (1990). Detecting Differential Item Functioning Using Logistic Regression Procedures. *Journal of Educational Measurement,* 27(4), 361-370. doi: 10.2307/1434855

Taylor, C. S., & Lee, Y. (2012). Gender DIF in Reading and Mathematics Tests With Mixed Item Formats. *Applied Measurement in Education,* 25(3), 246-280. doi: 10.1080/08957347.2012.687650

Tekwe, C. D., Carter, R. L., Ma, C.-X., Algina, J., Lucas, M. E., Roth, J., . . . Resnick, M. B. (2004). An Empirical Comparison of Statistical Models for Value-Added Assessment of School Performance. *Journal of Educational and Behavioral Statistics,* 29(1), 11-36. doi: 10.3102/10769986029001011

Tuksino, P., Kanjanawasee, S., & Pasiphol, C. (2010). A Quality Assessment of Science Instructional Management in Basic Education Schools: An Application of a Value-added Model and Differential Item Functioning. *The International Journal of Learning,* 17(5), 81-94.

Van de gaer, E., Grisay, A., Schulz, W., & Gebhardt, E. (2012). The Reference Group Effect: An Explanation of the Paradoxical Relationship Between Academic Achievement and Self-Confidence Across Countries. *Journal of Cross-Cultural Psychology,* 43(8), 1205-1228. doi: 10.1177/0022022111428083

Walker, C. M., Zhang, B., & Surber, J. (2008). Using a Multidimensional Differential Item Functioning Framework to Determine if Reading Ability Affects Student Performance in Mathematics. *Applied Measurement in Education,* 21(2), 162-181. doi: 10.1080/08957340801926201

Westers, P., & Kelderman, H. (1992). Examining differential item functioning due to item difficulty and alternative attractiveness. *Psychometrika,* 57(1), 107-118. doi: 10.1007/BF02294661

Wood, S. W. (2011). *Differential item functioning procedures for polytomous items when examinee sample sizes are small.* (PhD (Doctor of Philosophy), University of Iowa. Retrieved from http://ir.uiowa.edu/etd/1110

Xiaoming, X. (2010). How do we go about investigating test fairness? *Language Testing,* 27(2), 147-170. doi: 10.1177/0265532209349465

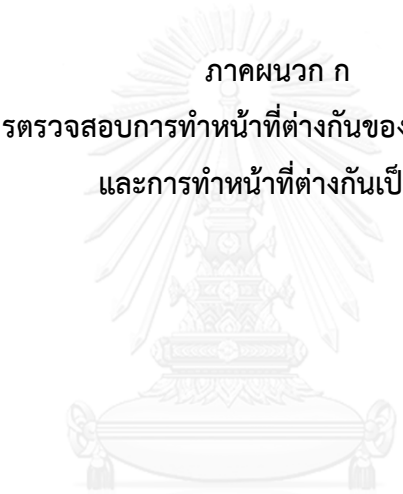
Yildirim, H. H., & Berberoglu, G. (2009). Judgmental and Statistical DIF Analyses of the PISA-2003 Mathematics Literacy Items. *International Journal of Testing,* 9(2), 108-121. doi: 10.1080/15305050902880736

Yildirim, H. H., & Yildirim, S. (2011). CORRELATES OF COMMUNALITIES AS MATCHING VARIABLES IN DIFFERENTIAL ITEM FUNCTIONING ANALYSES. *H. U. Journal of Education,* 40, 386-396.

ภาคผนวก

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาคผนวก ก
ข้อมูลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ และการทำหน้าที่ต่างกันเป็นขั้น



จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางแสดง ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ และการทำหน้าที่ต่างกันเป็นขั้น

Test	ข้อสอบ/สถิติทดสอบ		GENDER		EXTRA		ESCS		HEDRES		WEALTH	
No.2	DTF	τ^2	0.009	S	-0.039	S	0.042	S	0.013	S	-0.011	S
		z	0.391		-1.696		0.84		0.464		-0.25	
Poly	S465Q01	L-A LOR	-0.151		-0.447		1.472*		0.956*		1.137*	
		LOR Z	-0.812		1.862	-	6.571		5.005		4.512	
	Step1	CU-LOR	-0.171	S	-0.408	M	1.605	L	0.962	L	1.021	L
		Z	-0.921		-1.68		6.663		5.028		3.789	
	Step2	CU-LOR	-0.081	S	-0.574	M	1.11	L	0.937	L	1.443	L
		Z	-0.231		-1.273		3.071		2.582		3.815	
Dicho	S131Q02D	LOR Z	0.3009		-0.1715		0.2604		-0.9376		0.878	
	S131Q04D		-2.3568*		1.6233		-1.7966		-0.7342		-2.0082*	
	S415Q02		1.1173		0.6148		0.7874		0.7797		0.2315	
	S415Q07T		1.2926		0.2131		-1.4424		-0.7594		-0.5656	
	S415Q08T		-0.027		-0.728		1.948		2.1246*		1.6055	
	S428Q01		-1.0879		-0.7077		-0.1267		-0.5317		-0.2829	
	S428Q03		0.3304		-0.6907		1.592		0.9868		0.1925	
	S428Q05		-0.9732		-0.575		0.2234		1.3467		0.088	
	S438Q01T		0.0679		1.2557		-0.9069		0.3022		-0.9369	
	S438Q02		1.2265		-0.9072		-0.994		-2.3834*		-0.9027	
	S438Q03D		-0.1854		0.1279		-0.2427		1.7665		0.1396	
	S465Q02		-0.1856		-0.4023		0.5008		-0.5865		0.9318	
	S465Q04		-0.0884		0.6175		0.5043		-1.3928		-0.8541	
	S514Q02		0.003		-1.0505		-1.9114		-0.4132		0.6255	
	S514Q03		-0.1259		0.2453		0.8276		0.4107		0.4528	
	S514Q04		-0.738		0.6841		-0.1985		0.3909		0.1351	
No.3	DTF	τ^2	0.045	S	-0.036	S	0.042	S	0.004	S	0.085	M
		z	1.731		-1.636		0.032		0.211		1.441	
Poly	S465Q01	L-A LOR	-0.243		-0.461		1.291*		0.833*		1.084*	
		LOR Z	-1.286		-1.815		5.815		4.339		4.285	
	Step1	CU-LOR	-0.263	S	-0.524	M	1.188	L	0.793	L	0.978	L
		Z	-1.388		-2.029		5.022		4.087		3.61	
	Step2	CU-LOR	-0.189	S	-0.312	S	1.526	L	0.939	L	1.313	L
		Z	-0.603		-0.779		4.704		2.932		3.755	
	S498Q04	L-A LOR	0.19		-0.198		1.242*		0.991*		1.13*	
		LOR Z	1.005		-0.767		5.197		5.082		4.139	
	Step1	CU-LOR	0.218	S	-0.189	S	1.175	L	1.052	L	1.039	L
		Z	1.129		-0.71		4.716		5.228		3.667	
	Step2	CU-LOR	0.162	S	-0.207	S	1.31	L	0.928	L		L
		Z									1.224	

Test	ข้อสอบ/สถิติทดสอบ		GENDER	EXTRA	ESCS	HEDRES	WEALTH
Dicho	S131Q02D	LOR Z	1.9252	0.7062	-1.5326	-0.2005	-0.0815
	S131Q04D		0.2869	-0.1721	-1.5724	-1.2312	-2.2106*
	S256Q01		2.2067*	1.0051	0.2776	-0.2724	1.4978
	S326Q01		1.6752	-0.0134	-0.7407	0.4486	0.2583
	S326Q02		0.295	-0.6494	0.7685	2.5993*	0.5387
	S326Q03		-2.3861*	1.0493	0.7415	-1.1443	0.8002
	S326Q04T		0.7404	-1.108	-0.4412	-0.7738	-0.7298
	S413Q04T		-1.2565	-0.4031	-0.1079	0.137	-0.8535
	S413Q05		0.0217	-0.8048	-1.6041	-0.1152	-0.6879
	S413Q06		-1.4113	-1.5354	1.0007	0.929	0.7108
	S415Q02		-0.8368	-0.8283	0.2852	0.4564	1.3426
	S415Q07T		2.8522*	-0.0043	0.4349	-0.1227	-1.1261
	S415Q08T		0.6048	1.1765	-0.0888	-0.1273	-0.1345
	S425Q02		-1.4152	-1.6309	0.5417	1.469	0.5886
	S425Q03		-2.1692*	0.8967	-0.0144	1.2084	0.4593
	S425Q04		0.8845	0.1918	0.1311	0.5472	1.3071
	S425Q05		-0.4486	-0.3721	-0.0576	-1.7033	-0.6081
	S428Q01		-1.1187	-1.3028	1.6278	0.3835	1.8789
	S428Q03		-1.6621	0.954	0.4086	-0.5779	-0.95
	S428Q05		-0.9252	-0.1995	2.9534*	2.3439*	0.5806
	S438Q01T		0.3377	1.0695	-0.4096	-2.0994*	-0.9556
	S438Q02		-0.1976	-0.1618	-1.7444	-0.2583	-1.3246
	S438Q03D		1.8899	0.5773	1.4744	0.3707	0.4519
	S465Q02		-0.0631	0.0079	0.7377	0.3865	1.029
	S465Q04		-1.2827	-0.0958	-0.539	-0.5932	1.4284
	S478Q01		0.1414	1.6518	0.8287	-1.57	-0.6584
	S478Q02T		-1.1049	0.9526	-0.8727	-0.787	-0.6578
	S478Q03T		1.1119	0.0016	-0.835	0.591	0.4856
	S498Q02T		-0.3365	0.9295	-0.1714	0.2302	-1.9249
	S498Q03		0.1611	-1.3005	-2.5183*	-1.2174	-2.3073*
	S514Q02		0.9537	0.9393	1.2654	1.4905	1.9617
	S514Q03		0.1608	-1.18	-1.3473	-0.6447	-0.1778
	S514Q04		0.6138	-0.8422	1.9092	1.7781	1.2139

Test	ข้อสอบ/สถิติทดสอบ		GENDER		EXTRA		ESCS		HEDRES		WEALTH	
NO. 4	DTF	τ^2	0.016	S	-0.013	S	-0.007	S	-0.026	S	-0.021	S
		z	0.64		-0.464		-0.241		-2.000		-0.808	
Poly	S519Q01	L-A LOR	0.495*		-0.115		1.131*		0.989*		0.916*	
		LOR Z	2.647		-0.494		5.164		5.289		4.000	
	Step1	CU-LOR	0.486	M	-0.145	S	1.225	L	1.043	L	1.056	L
		Z	2.554		-0.592		4.982		5.347		4.128	
	Step2	CU-LOR	0.515	M	-0.06	S		L		L		L
		Z					0.975		0.882		0.677	
Dicho	S269Q01	LOR Z	-1.2594		0.0265		1.1254		-0.6723		0.7897	
	S269Q03D		-1.4871		-0.8767		1.3602		1.2858		1.3983	
	S269Q04T		-1.3216		0.5517		-0.1043		0.6244		0.084	
	S408Q01		0.2676		0.2375		0.4991		0.5489		-0.0032	
	S408Q03		1.6691		1.5548		-1.2531		-0.664		0.0924	
	S408Q04T		-0.0163		0.2629		-0.9195		-0.8716		0.1979	
	S408Q05		-0.9216		0.3529		-0.7135		-1.1674		-0.1256	
	S466Q01T		1.7414		-0.7401		-0.2351		-0.9012		-1.8887	
	S466Q05		-0.3149		-1.1329		-0.1925		1.0786		0.5264	
	S466Q07T		0.8611		2.2947*		1.4308		-0.0184		-0.3278	
	S519Q02T		-0.8764		-0.9891		-1.4469		0.0836		0.1305	
	S519Q03		-0.2426		-1.4009		0.5451		0.6335		-0.1654	
	S521Q02		1.1033		0.62		-0.9358		-0.874		-1.2142	
	S521Q06		1.3953		-0.625		1.5566		0.2397		2.0386*	
	S527Q01T		0.7461		-0.6846		1.1213		0.1561		0.0249	
	S527Q03T		-1.4967		-0.4211		-0.5585		1.0058		-0.1635	
	S527Q04T		-0.0705		0.5527		-0.5966		-0.1777		-0.9163	
No. 7	DTF	τ^2	0.076	M	0.027	S	0.006	S	-0.031	S	-0.026	S
		z	1.652		0.529		0.158		-2.067		-0.703	
Poly	S498Q04	L-A LOR	0.558*		-0.263		1.012*		0.821*		1.113*	
		LOR Z	3.1		-1.04		4.419		4.344		3.838	
	Step1	CU-LOR	0.613	L	-0.238	S	0.958	L	0.84	L	1.068	L
		Z	3.218		-0.879		3.865		4.133		3.378	
	Step2	CU-LOR	0.505	M	-0.285	S	1.062	L	0.802	L	1.155	L
		Z	2.697		-1.105		4.554		4.136		3.952	

Test	ข้อสอบ/สถิติทดสอบ		GENDER		EXTRA		ESCS		HEDRES		WEALTH	
Dicho	S256Q01	LOR Z	2.187*		0.5974		0.2258		0.0805		-1.8651	
	S326Q01		2.0629*		-0.8114		-0.5825		0.223		-1.1418	
	S326Q02		0.8645		-2.5082*		0.0221		-0.1419		1.2193	
	S326Q03		-1.7319		0.3505		0.1486		0.1019		0.5337	
	S326Q04T		-1.1458		-1.1732		0.695		0.2114		-0.1843	
	S413Q04T		0.5251		0.3171		-2.6346*		-0.7187		-0.5789	
	S413Q05		0.0099		0.4345		-0.4185		-1.3495		-0.671	
	S413Q06		-1.08		-0.9886		1.0219		0.4162		-0.0291	
	S425Q02		-0.733		1.5363		0.39		-0.2275		1.3139	
	S425Q03		0.368		0.2412		0.9672		0.8116		0.3383	
	S425Q04		2.3127*		1.1729		0.6414		0.0347		0.6466	
	S425Q05		-1.8653		0.1946		-0.8628		1.0923		0.8526	
	S478Q01		-0.3647		0.1889		-2.0844*		-1.637		-1.561	
	S478Q02T		1.0855		-1.0569		2.479*		2.123*		0.3897	
	S478Q03T		-0.9673		1.9884		1.2526		-0.9255		0.111	
	S498Q02T		-0.6026		-0.8972		-0.3629		0.2995		-0.1057	
	S498Q03		2.187*		0.5974		0.2258		0.0805		-1.8651	
No. 8	DTF	τ^2	0.118	M	-0.028	S	0.128	M	0.012	S	0	S
		z	1.844		-1.037		1.561		0.429		0	
Poly	S465Q01	L-A LOR	0.062		-0.14		0.773*		0.697*		1.186*	
		LOR Z	0.32		-0.571		3.498		3.556		4.492	
	Step1	CU-LOR	0.059	S	-0.044	S	0.639	L	0.603	L	1.136	L
		Z	0.306		-0.177		2.843		3.085		4.065	
	Step2	CU-LOR	0.074	S	-0.417	S	1.225	L	0.66	L	1.329	L
		Z	0.194		-0.964		3.187		2.724		3.207	
Dicho	S131Q02D	LOR Z	3.3668*		-0.74		-0.329		0.5449		0.0389	
	S131Q04D		1.69		1.071		-2.053*		-1.3412		-1.0467	
	S415Q02		-0.8429		-1.0055		0.8264		0.3697		0.2073	
	S415Q07T		0.4079		-0.2318		-1.5958		-1.4626		-0.7005	
	S415Q08T		-0.5855		-0.3467		1.6111		1.4753		-0.055	
	S428Q01		-2.3461*		-1.2281		0.8148		0.8409		0.7319	
	S428Q03		-0.0573		-0.08		1.2247		2.3397*		1.6539	
	S428Q05		-1.9028		1.2679		0.4718		-0.7738		-1.1878	
	S438Q01T		-0.216		0.3946		-1.02		-0.5613		-1.4658	
	S438Q02		1.3529		-0.6684		0.9024		0.4692		1.6043	
	S438Q03D		1.5564		0.3595		1.2346		-0.5733		0.3793	
	S465Q02		-2.3975*		-0.5299		-0.3279		-1.79		-0.5818	
	S465Q04		1.9359		-0.0686		-0.8669		-0.6938		0.5219	
	S514Q02		-1.4592		1.0532		0.542		1.4712		-0.2311	
	S514Q03		0.3639		1.2314		-2.0619*		-0.2267		1.4593	
	S514Q04		-0.8777		0.4454		0.453		-0.2462		-1.3889	

Test	ข้อสอบ/สถิติทดสอบ		GENDER		EXTRA		ESCS		HEDRES		WEALTH	
No. 9	DTF	τ^2	0.1	M	0.003	S	0.035	S	0.01	S	-0.021	S
		z	1.923		0.088		0.875		0.435		-0.7	
Poly	S519Q01	L-A LOR	0.374*		-0.577*		0.862*		0.703*		0.595*	
		LOR Z	2.089		-2.455		4.205		3.949		2.38	
	Step1	CU-LOR	0.408	M	-0.443	M	1.033	L	0.704	L	0.456	L
		Z	2.169		-1.703		4.377		3.681		1.63	
	Step2	CU-LOR	0.317	M	-0.772	M	0.622	L	0.702	L	0.793	L
		Z	1.275		-2.374		2.396		2.913		2.598	
Dicho	S269Q01	LOR Z	-2.9057*		-0.9154		0.6125		1.318		-0.0008	
	S269Q03D		-0.8321		-0.6353		2.1238*		0.7614		0.2749	
	S269Q04T		-0.9218		0.1657		0.1723		-0.1888		0.6842	
	S408Q01		1.6935		0.1211		0.5119		0.8958		-0.0081	
	S408Q03		1.1323		0.375		-1.2077		-1.0449		-0.7642	
	S408Q04T		-0.7222		-0.964		-2.328*		-1.7627		-0.439	
	S408Q05		-1.4179		0.1833		-0.4542		-0.6124		0.9169	
	S466Q01T		1.0045		-0.8552		0.7799		0.8149		0.8804	
	S466Q05		0.8411		0.9323		-1.2185		-0.8496		0.0945	
	S466Q07T		2.5098*		2.5064*		0.1294		-1.7533		-1.7511	
	S519Q02T		0.3598		0.6755		-0.6412		0.2179		-0.9459	
	S519Q03		1.1739		0.1244		1.7602		1.0548		2.1395*	
	S521Q02		-1.5834		0.2664		-1.1375		-0.711		-0.3315	
	S521Q06		2.8046*		-0.9906		2.1836*		1.9817		1.2699	
	S527Q01T		-0.3611		-0.6462		-0.5156		-0.9102		0.6743	
	S527Q03T		0.643		0.893		-0.0054		0.7838		-0.4239	
	S527Q04T		-3.2958*		-1.5268		0.403		0.3781		-1.2313	
No. 10	DTF	τ^2	0.027	S	0.056	S	0.084	M	0.022	S	0.057	S
		z	1.174		1.302		1.867		1		1.213	
Poly	S465Q01	L-A LOR	0.247		-0.622*		0.707*		0.52*		0.439	
		LOR Z	1.171		-2.365		3.115		2.574		1.535	
	Step1	CU-LOR	0.415	M	-0.351	S	0.542	M	0.46	M	0.354	S
		Z	2.01		-1.323		2.387		2.306		1.239	
	Step2	CU-LOR	-0.554	M	-2.573	L	1.471	L	0.827	L	0.799	L
		Z	-1.237		-2.46		3.17		1.799		1.5	
	S519Q01	L-A LOR	0.351		-0.296		1.031*		0.821*		1.278*	
		LOR Z	1.961		-1.265		4.818		4.561		4.681	
	Step1	CU-LOR	0.454	M	-0.221	S	1.188	L	0.919	L	1.366	L
		Z	2.334		-0.817		4.413		4.505		3.648	
	Step2	CU-LOR	0.188	S	-0.387	S	0.846	L	0.677	L	1.191	L
		Z	0.758		-1.288		3.292		2.794		4.012	

Test	ข้อสอบ/สถิติทดสอบ		GENDER	EXTRA	ESCS	HEDRES	WEALTH
Dicho	S131Q02D	LOR Z	1.3163	0.6584	0.6473	-0.0187	-0.8422
	S131Q04D		0.4143	-0.7761	1.2436	-0.0419	0.8741
	S269Q01		-2.1529*	-0.5021	-0.9399	0.8257	0.4964
	S269Q03D		-1.1628	0.1463	0.7849	0.955	-0.275
	S269Q04T		-1.5376	-0.6221	0.5967	-0.0495	-0.3201
	S408Q01		-1.1482	1.6169	1.3578	2.1904*	1.6402
	S408Q03		-0.4077	0.4163	-1.3918	-2.8682	-0.1711
	S408Q04T		1.1093	-1.2031	-0.4385	0.1011	-1.5805
	S408Q05		-2.3457*	0.9767	-1.1993	-0.0803	-1.0391
	S415Q02		0.6441	-1.4779	-0.2214	-2.0068*	0.349
	S415Q07T		1.6091	0.5232	0.0807	-0.5028	-0.0688
	S415Q08T		-0.5119	-1.1833	2.5836*	0.9719	1.6166
	S428Q01		-0.9767	0.003	1.6108	1.5124	1.1728
	S428Q03		0.2066	-0.354	0.1381	-0.3768	-1.2464
	S428Q05		0.838	-0.0268	0.8186	0.7757	0.4344
	S438Q01T		-0.9833	-0.0087	-2.6425*	-0.8468	-0.3767
	S438Q02		2.5461*	-2.0374*	1.2352	-0.1457	0.2218
	S438Q03D		-0.2311	-0.3851	1.7995	0.4245	2.2772*
	S465Q02		0.8275	1.0564	1.1549	1.7619	-0.2021
	S465Q04		-0.4786	1.5475	-2.4127*	-1.1733	-0.4396
	S466Q01T		0.5821	0.3555	-0.5249	-2.1349*	-0.1577
	S466Q05		-0.4637	2.2126*	-0.8945	-0.4601	-0.9484
	S466Q07T		2.6921*	0.9436	-0.4489	-0.614	-1.4038
	S514Q02		0.4531	-1.8614	-1.1123	0.2487	0.0968
	S514Q03		0.7529	0.2231	-0.3144	-1.5637	-1.8948
	S514Q04		1.3029	-2.3173*	2.2869*	2.2958*	2.1953*
	S519Q02T		-0.1072	-0.6963	-0.0428	1.3677	-0.09
	S519Q03		0.1582	-1.019	1.8303	1.7424	2.29*
	S521Q02		0.4858	1.8766	-2.0678*	-1.3684	-1.3533
	S521Q06		0.2286	1.4478	1.5395	0.7361	0.1717
	S527Q01T		-1.3404	0.4628	-0.8504	0.1659	0.5146
	S527Q03T		-0.882	-1.414	-0.9165	0.403	0.3906
	S527Q04T		-1.8751	0.6479	-1.5092	-1.1717	-1.1887

Test	ข้อสอบ/สถิติทดสอบ		GENDER		EXTRA		ESCS		HEDRES		WEALTH	
NO.12	DTF	τ^2	0.123	M	0.017	S	0.059	S	0.012	S	0.089	M
		z	2.733		0.459		1.553		0.6		1.745	
Poly	S498Q04	L-A LOR	0.609*		-0.664*		1.069*		0.922*		0.721*	
		LOR Z	3.346		-2.604		4.53		4.878		2.641	
	Step1	CU-LOR	0.63	L	-0.758	L	1.196	L	1.059	L	0.848	L
		Z	3.202		-2.706		4.382		5.044		2.693	
	Step2	CU-LOR	0.589	M	-0.585	M	0.966	L	0.806	L	0.618	L
		Z	3.146		-2.263		4.155		4.221		2.28	
	S519Q01	L-A LOR	0.486*		-0.77*		1.301*		1.082*		1.459*	
		LOR Z	2.505		-2.939		6.051		5.492		5.722	
	Step1	CU-LOR	0.466	M	-0.879	L	1.284	L	1.043	L	1.445	L
		Z	2.395		-3.266		5.62		5.263		5.076	
	Step2	CU-LOR	0.55	M	-0.535	L	1.347	L	1.208	L	1.488	L
		Z	1.554		-1.316		3.937		3.267		4.175	
Dicho	S256Q01	LOR Z	0.5781		0.0428		-1.1066		-0.5659		-2.4621*	
	S269Q01		-0.6588		0.3401		-0.2734		1.5929		0.5344	
	S269Q03D		0.6671		-1.3299		-1.3503		0.2653		-0.0256	
	S269Q04T		-2.9081*		-1.5129		0.8218		0.1127		1.1365	
	S326Q01		2.9335*		-1.1378		1.2528		1.8289		2.3573*	
	S326Q02		2.2627*		0.9436		-0.4782		0.3179		0.7885	
	S326Q03		-0.5932		-0.6134		1.8244		2.3311*		1.328	
	S326Q04T		-1.3839		-1.2544		2.3821*		2.422*		1.7078	
	S408Q01		0.4564		-0.087		-0.0907		0.0182		0.3602	
	S408Q03		-0.3596		-0.1129		-2.4999*		-0.924		-1.983	
	S408Q04T		2.3153*		1.7096		-0.8448		-0.8187		-1.2743	
	S408Q05		1.4571		1.2355		0.4244		0.6433		1.1065	
	S413Q04T		-2.405*		-0.8622		0.1741		0.5608		0.1215	
	S413Q05		0.8625		1.5405		-0.076		0.8528		-0.8506	
	S413Q06		-1.8144		-1.2222		2.9778*		0.1479		1.7481	
	S425Q02		0.3903		-1.8663		1.6163		0.5685		1.6192	
	S425Q03		1.2659		0.2803		0.0379		-0.7353		-0.4968	

Test	ข้อสอบ/สถิติทดสอบ		GENDER	EXTRA	ESCS	HEDRES	WEALTH
	S425Q04		1.0318	0.6861	-0.0089	-0.2943	-1.1775
	S425Q05		0.0113	1.2201	0.4357	0.023	-0.3267
	S466Q01T		-0.149	0.5027	0.683	-0.5464	0.592
	S466Q05		-3.2454*	-0.3072	1.476	-0.1273	1.2871
	S466Q07T		3.014*	1.0601	0.1654	0.5504	-1.6839
	S478Q01		-0.5087	-0.3335	-1.254	-0.1442	-1.0055
	S478Q02T		-1.8086	-0.5772	1.7143	1.0021	0.8003
	S478Q03T		1.6202	0.237	-0.9997	-1.6774	-0.1346
	S498Q02T		-0.5191	0.4289	-0.5032	-0.9794	-0.1995
	S498Q03		-0.0517	0.4559	-0.1804	-0.5817	0.4062
	S519Q02T		-1.0042	1.5781	-2.6614*	-1.3526	-1.9509
	S519Q03		1.9836	-0.4908	-0.1924	0.7814	-0.8724
	S521Q02		-1.8763	1.0374	-1.4863	-1.1012	0.1406
	S521Q06		-1.3663	-0.4974	-0.7446	-0.6485	-1.2459
	S527Q01T		-0.5874	-1.5919	-0.7906	0.2919	-0.4204
	S527Q03T		0.5781	0.0428	-1.1066	-0.5659	-2.4621*
	S527Q04T		-0.6588	0.3401	-0.2734	1.5929	0.5344
No. 13	**DTF**	τ^2	0.025 **S**	0.001 **S**	0.03 **S**	0.08 **M**	-0.013 **S**
		z	0.926	0.029	0.769	1.633	-0.406
Poly	S498Q04	**L-A LOR**	0.2	-0.577*	1.059*	0.861*	0.806*
		LOR Z	1.111	-2.281	4.34	4.438	2.952
	Step1	**CU-LOR**	1.555 **L**	-0.647 **L**	1.05 **L**	0.921 **L**	0.702 **L**
		Z	0.802	-2.335	3.91	4.338	2.398
	Step2	**CU-LOR**	0.241 S	-0.517 **M**	1.067 **L**	0.809 **L**	0.899 **L**
		Z	1.305	-2.021	4.378	4.132	3.243
Dicho	S256Q01	**LOR Z**	0.942	0.3673	-1.1733	-2.4264*	0.4432
	S326Q01		1.2428	-0.2959	-1.029	-2.2859*	-1.1068
	S326Q02		-1.4501	-0.5122	2.0499*	2.6371*	2.0272*
	S326Q03		-0.4422	-2.1146*	-0.0355	-0.2171	0.0883
	S326Q04T		0.7111	-1.5841	0.0755	-0.1472	0.8019
	S413Q04T		-0.4795	0.7568	-2.1021*	-0.4152	-0.8476
	S413Q05		-1.2646	-1.0753	0.6997	-0.6909	0.5908
	S413Q06		0.3535	-1.6486	2.7137*	2.3434*	1.2642

Test	ข้อสอบ/สถิติทดสอบ		GENDER	EXTRA	ESCS	HEDRES	WEALTH
	S425Q02		-0.5773	-0.0971	0.0731	0.0739	0.3705
	S425Q03		-0.6703	-0.2972	1.3664	1.8641	0.4893
	S425Q04		1.5791	0.7427	-0.6681	1.1775	-0.9343
	S425Q05		-1.4806	-0.2869	-0.6071	0.0116	-1.292
	S478Q01		0.6248	1.2856	-0.5378	0.3354	0.125
	S478Q02T		-2.137*	0.9411	-0.3412	0.1987	-1.2953
	S478Q03T		2.5235*	1.2895	0.5952	-0.0763	0.5249
	S498Q02T		-0.7384	1.3818	-0.8268	-1.5269	-1.2081
	S498Q03		1.6192	0.3852	-0.171	-0.8212	0.5869

หมายเหตุ

Poly คือ polytomous items

Dicho คือ Dichotomous items

เกณฑ์พิจารณา DTF: $\tau^2 < 0.07$ small DIF effect variance

$0.07 < \tau^2 < 0.14$ medium DIF effect variance

$\tau^2 > 0.014$ large DIF effect variance

DIF : Dichotomous and polytomous items

LOR Z >2.00 presence of DIF (Reference group) and

LOR Z<-2.00 presence of DIF (Focal group)

DSF : Differential step functioning for polytomous items

| CU-LOR | < .4 small effect

.4 < | CU-LOR| < .6 medium effect

| CU-LOR| > .6 large effect

ภาคผนวก ข
ข้อสอบที่ถูกตัดแยกรายฉบับและภาพรวม

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางแสดง ข้อสอบที่ถูกตัดแยกรายฉบับและภาพรวม

ข้อสอบ	แบบสอบฉบับที่									
	2	3	4	7	8	9	10	12	13	รวมทั้งหมด
S131Q02D	✗	✗			✗		✗			✗
S131Q04D	✓	✓			✓		✓			✓
S256Q01		✓		✓				✓	✓	✓
S269Q01			✓			✓	✓	✓		✓
S269Q03D			✓			✓	✓	✓		✓
S269Q04T			✗			✗	✗	✗		✗
S326Q01		✗		✗				✗	✗	✗
S326Q02		✓		✓				✓	✓	✓
S326Q03		✓		✓				✓	✓	✓
S326Q04T		✗		✗				✗	✗	✗
S408Q01			✓			✓	✓	✓		✓
S408Q03			✓			✓	✓	✓		✓
S408Q04T			✓			✓	✓	✓		✓
S408Q05			✓			✓	✓	✓		✓
S413Q04T		✓		✓				✓	✓	✓
S413Q05		✓		✓				✓	✓	✓
S413Q06		✓		✓				✓	✓	✓
S415Q02	✓	✓			✓		✓			✓
S415Q07T	✓	✓			✓		✓			✓
S415Q08T	✓	✓			✓		✓			✓
S425Q02		✓		✓				✓	✓	✓
S425Q03		✓		✓				✓	✓	✓
S425Q04		✓		✓				✓	✓	✓

ข้อสอบ	แบบสอบฉบับที่									
	2	3	4	7	8	9	10	12	13	รวมทั้งหมด
S425Q05		✓		✓				✓	✓	✓
S428Q01	✓	✓			✓		✓			✓
S428Q03	✓	✓			✓		✓			✓
S428Q05	✓	✓			✓		✓			✓
S438Q01T	✓	✓			✓		✓			✓
S438Q02	✓	✓			✓		✓			✓
S438Q03D	✓	✓			✓		✓			✓
S465Q01	✗	✗			✗		✗			✗
S465Q02	✓	✓			✓		✓			✓
S465Q04	✓	✓			✓		✓			✓
S466Q01T			✓			✓	✓	✓		✓
S466Q05			✓			✓	✓	✓		✓
S466Q07T			✗			✗	✗	✗		✗
S478Q01		✓		✓				✓	✓	✓
S478Q02T		✓		✓				✓	✓	✓
S478Q03T		✓		✓				✓	✓	✓
S498Q02T		✓		✓				✓	✓	✓
S498Q03		✓		✓				✓	✓	✓
S498Q04		✗		✗				✗	✗	✗
S514Q02	✓	✓			✓		✓			✓
S514Q03	✓	✓			✓		✓			✓
S514Q04	✓	✓			✓		✓			✓
S519Q01			✗			✗	✗	✗		✗
S519Q02T			✓			✓	✓	✓		✓
S519Q03			✓			✓	✓	✓		✓
S521Q02			✓			✓	✓	✓		✓

ข้อสอบ	แบบสอบฉบับที่									
	2	3	4	7	8	9	10	12	13	รวมทั้งหมด
S521Q06			✓			✓	✓	✓		✓
S527Q01T			✓			✓	✓	✓		✓
S527Q03T			✓			✓	✓	✓		✓
S527Q04T			✓			✓	✓	✓		✓
จำนวนข้อทั้งหมด	17	35	18	18	17	18	35	36	18	53
ข้อที่ถูกตัด	2	5	3	3	2	3	5	6	3	8
เหลือ	15	30	15	15	15	15	30	30	15	45

หมายเหตุ

✗ หมายถึง ข้อสอบที่ถูกตัดทิ้ง

✓ หมายถึง ข้อสอบที่คงไว้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาคผนวก ค
ผลการวิเคราะห์โมเดลมูลค่าเพิ่มพหุระดับ 4 โมเดล

ผลการวิเคราะห์โมเดลที่ 1: โมเดลไม่มีเงื่อนไข Undetected DIF-DSF-DTF

Program: HLM 7 Hierarchical Linear and Nonlinear Modeling
Authors: Stephen Raudenbush, Tony Bryk, & Richard Congdon
Publisher: Scientific Software International, Inc. (c) 2010
techsupport@ssicentral.com
www.ssicentral.com

Module: HLM2.EXE (7.01.21202.1001)
Date: 30 June 2014, Monday
Time: 12:0:41

Specifications for this HLM2 run

Problem Title: no title

The data source for this run = VAmodel.mdm
The command file for this run = C:\Users\NINJANAT\Desktop\HLM value-added\Analysis VA before cut items\VA null model.hlm
Output file name = C:\Users\NINJANAT\Desktop\HLM value-added\Analysis VA before cut items\hlm2.html
The maximum number of level-1 units = 2358
The maximum number of level-2 units = 223
The maximum number of iterations = 100

Method of estimation: restricted maximum likelihood

The outcome variable is SCIACH

Summary of the model specified

Level-1 Model

$$SCIACH_{mj} = \psi_{0j} + e_{mj}$$

Level-2 Model

$$\psi_{0j} = \gamma_{00} + u_{0j}$$

Mixed Model

$$SCIACH_{mj} = \gamma_{00} + u_{0j} + e_{mj}$$

Final Results - Iteration 4

Iterations stopped due to small change in likelihood function

σ^2_ε = 4676.74019

τ
INTRCPT1,ψ_0 2180.53120

Random level-1 coefficient	Reliability estimate
INTRCPT1,ψ_0	0.778

The value of the log-likelihood function at iteration 4 = -1.348995E+004

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. $d.f.$	p-value
For INTRCPT1, ψ_0					
INTRCPT2, γ_{00}	492.673615	3.545497	138.958	222	<0.001

Final estimation of fixed effects (with robust standard errors)

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. $d.f.$	p-value
For INTRCPT1, ψ_0					
INTRCPT2, γ_{00}	492.673615	3.537023	139.290	222	<0.001

Final estimation of variance components

Random Effect	Standard Deviation	Variance Component	$d.f.$	χ^2	p-value
INTRCPT1, u_0	46.69616	2180.53120	222	1549.95905	<0.001
level-1, e	68.38670	4676.74019			

Statistics for current covariance components model

Deviance = 26979.904720
Number of estimated parameters = 2

ผลการวิเคราะห์โมเดลที่ 2: โมเดล Undetected DIF-DSF-DTF

Program:	HLM 7 Hierarchical Linear and Nonlinear Modeling
Authors:	Stephen Raudenbush, Tony Bryk, & Richard Congdon
Publisher:	Scientific Software International, Inc. (c) 2010
	techsupport@ssicentral.com
	www.ssicentral.com

Module:	HLM2.EXE (7.01.21202.1001)
Date:	30 June 2014, Monday
Time:	12:3:24

Specifications for this HLM2 run

Problem Title: no title

The data source for this run = VAmodel.mdm
The command file for this run = C:\Users\NINJANAT\Desktop\HLM value-added\Analysis VA before cut items\VA model before cut items.hlm
Output file name = C:\Users\NINJANAT\Desktop\HLM value-added\Analysis VA before cut items\hlm2.html
The maximum number of level-1 units = 2358
The maximum number of level-2 units = 223
The maximum number of iterations = 100

Method of estimation: restricted maximum likelihood

The outcome variable is SCIACH

Summary of the model specified

Level-1 Model

$$SCIACH_{mj} = \psi_{0j} + \psi_{1j}*(GENDER_{mj}) + \psi_{2j}*(EXTRA_{mj}) + \psi_{3j}*(ESCS_{mj}) + \psi_{4j}*(HEDRES_{mj}) + \psi_{5j}*(WEALTH_{mj}) + \psi_{6j}*(ICTRES_{mj}) + e_{mj}$$

Level-2 Model

$$\psi_{0j} = \gamma_{00} + \gamma_{01}*(PUBLIC_j) + \gamma_{02}*(TCSHORT_j) + \gamma_{03}*(FUNGOV_j) + \gamma_{04}*(SCMATEDU_j) + \gamma_{05}*(STRATIO_j) + \gamma_{06}*(RESPCURR_j) + u_{0j}$$

$$\psi_{1j} = \gamma_{10}$$

$$\psi_{2j} = \gamma_{20}$$

$\psi_{3j} = \gamma_{30}$
$\psi_{4j} = \gamma_{40}$
$\psi_{5j} = \gamma_{50}$
$\psi_{6j} = \gamma_{60}$

ESCS HEDRES WEALTH ICTRES have been centered around the grand mean.

TCSHORT FUNGOV SCMATEDU STRATIO RESPCURR have been centered around the grand mean.

Mixed Model

$$SCIACH_{mj} = \gamma_{00} + \gamma_{01}*PUBLIC_j + \gamma_{02}*TCSHORT_j + \gamma_{03}*FUNGOV_j + \gamma_{04}*SCMATEDU_j + \gamma_{05}*STRATIO_j + \gamma_{06}*RESPCURR_j + \gamma_{10}*GENDER_{mj} + \gamma_{20}*EXTRA_{mj} + \gamma_{30}*ESCS_{mj} + \gamma_{40}*HEDRES_{mj} + \gamma_{50}*WEALTH_{mj} + \gamma_{60}*ICTRES_{mj} + u_{0j} + e_{mj}$$

Final Results - Iteration 7

Iterations stopped due to small change in likelihood function

$\sigma^2_\varepsilon = 4573.91458$

τ

INTRCPT1, ψ_0	874.87160

Random level-1 coefficient	Reliability estimate
INTRCPT1, ψ_0	0.614

The value of the log-likelihood function at iteration 7 = -1.337056E+004

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *d.f.*	*p*-value
For INTRCPT1, ψ_0					
INTRCPT2, γ_{00}	465.761658	8.014537	58.115	216	<0.001
PUBLIC, γ_{01}	36.595242	8.110209	4.512	216	<0.001
TCSHORT, γ_{02}	3.543837	2.807105	1.262	216	0.208
FUNGOV, γ_{03}	-0.413713	0.118185	-3.501	216	<0.001
SCMATEDU, γ_{04}	12.413339	3.105312	3.997	216	<0.001
STRATIO, γ_{05}	-0.197582	0.379741	-0.520	216	0.603

RESPCURR, γ_{06}	2.185643	3.259576	0.671	216	0.503
For GENDER slope, ψ_1					
INTRCPT2, γ_{10}	1.253225	3.080387	0.407	2129	0.684
For EXTRA slope, ψ_2					
INTRCPT2, γ_{20}	-4.105364	3.075471	-1.335	2129	0.182
For ESCS slope, ψ_3					
INTRCPT2, γ_{30}	13.183396	2.099975	6.278	2129	<0.001
For HEDRES slope, ψ_4					
INTRCPT2, γ_{40}	10.708983	2.327781	4.601	2129	<0.001
For WEALTH slope, ψ_5					
INTRCPT2, γ_{50}	-16.088422	2.799750	-5.746	2129	<0.001
For ICTRES slope, ψ_6					
INTRCPT2, γ_{60}	5.036598	2.030588	2.480	2129	0.013

Final estimation of fixed effects
(with robust standard errors)

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *d.f.*	*p*-value
For INTRCPT1, ψ_0					
INTRCPT2, γ_{00}	465.761658	8.482754	54.907	216	<0.001
PUBLIC, γ_{01}	36.595242	8.463717	4.324	216	<0.001
TCSHORT, γ_{02}	3.543837	2.713442	1.306	216	0.193
FUNGOV, γ_{03}	-0.413713	0.107688	-3.842	216	<0.001
SCMATEDU, γ_{04}	12.413339	3.257705	3.810	216	<0.001
STRATIO, γ_{05}	-0.197582	0.394244	-0.501	216	0.617
RESPCURR, γ_{06}	2.185643	3.072819	0.711	216	0.478
For GENDER slope, ψ_1					
INTRCPT2, γ_{10}	1.253225	3.428090	0.366	2129	0.715
For EXTRA slope, ψ_2					
INTRCPT2, γ_{20}	-4.105364	3.192702	-1.286	2129	0.199
For ESCS slope, ψ_3					
INTRCPT2, γ_{30}	13.183396	2.138834	6.164	2129	<0.001
For HEDRES slope, ψ_4					
INTRCPT2, γ_{40}	10.708983	2.548756	4.202	2129	<0.001
For WEALTH slope, ψ_5					
INTRCPT2, γ_{50}	-16.088422	2.547443	-6.316	2129	<0.001
For ICTRES slope, ψ_6					
INTRCPT2, γ_{60}	5.036598	2.115399	2.381	2129	0.017

Final estimation of variance components

Random Effect	Standard Deviation	Variance Component	*d.f.*	χ^2	*p*-value
INTRCPT1, u_0	29.57823	874.87160	216	684.76089	<0.001
level-1, e	67.63072	4573.91458			

Statistics for current covariance components model

Deviance = 26741.118290
Number of estimated parameters = 2

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ผลการวิเคราะห์โมเดลที่ 3: โมเดลไม่มีเงื่อนไข Detected DIF-DSF-DTF

Program: HLM 7 Hierarchical Linear and Nonlinear Modeling
Authors: Stephen Raudenbush, Tony Bryk, & Richard Congdon
Publisher: Scientific Software International, Inc. (c) 2010
techsupport@ssicentral.com
www.ssicentral.com

Module: HLM2.EXE (7.01.21202.1001)
Date: 30 June 2014, Monday
Time: 9:52:36

Specifications for this HLM2 run

Problem Title: no title

The data source for this run = VA null model
The command file for this run = C:\Users\NINJANAT\Desktop\Analysis VA after cut items\VA null model after cut items.hlm
Output file name = C:\Users\NINJANAT\Desktop\Analysis VA after cut items\hlm2.html
The maximum number of level-1 units = 2358
The maximum number of level-2 units = 223
The maximum number of iterations = 100

Method of estimation: restricted maximum likelihood

The outcome variable is SCIACH

Summary of the model specified

Level-1 Model

$$SCIACH_{mj} = \psi_{0j} + e_{mj}$$

Level-2 Model

$$\psi_{0j} = \gamma_{00} + u_{0j}$$

Mixed Model

$$SCIACH_{mj} = \gamma_{00} + u_{0j} + e_{mj}$$

Final Results - Iteration 4

Iterations stopped due to small change in likelihood function

σ^2_ε = 4627.52810

τ
INTRCPT1,ψ_0 2002.35002

Random level-1 coefficient	Reliability estimate
INTRCPT1,ψ_0	0.766

The value of the log-likelihood function at iteration 4 = -1.347109E+004

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *d.f.*	*p*-value
For INTRCPT1, ψ_0					
INTRCPT2, γ_{00}	492.245690	3.422905	143.809	222	<0.001

Final estimation of fixed effects
(with robust standard errors)

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *d.f.*	*p*-value
For INTRCPT1, ψ_0					
INTRCPT2, γ_{00}	492.245690	3.415487	144.122	222	<0.001

Final estimation of variance components

Random Effect	Standard Deviation	Variance Component	*d.f.*	χ^2	*p*-value
INTRCPT1, u_0	44.74763	2002.35002	222	1455.20082	<0.001
level-1, e	68.02594	4627.52810			

Statistics for current covariance components model

Deviance = 26942.189574
Number of estimated parameters = 2

ผลการวิเคราะห์โมเดลที่ 4: โมเดล Detected DIF-DSF-DTF

Program:	HLM 7 Hierarchical Linear and Nonlinear Modeling
Authors:	Stephen Raudenbush, Tony Bryk, & Richard Congdon
Publisher:	Scientific Software International, Inc. (c) 2010
	techsupport@ssicentral.com
	www.ssicentral.com

Module:	HLM2.EXE (7.01.21202.1001)
Date:	28 June 2014, Saturday
Time:	19:37: 3

Specifications for this HLM2 run

Problem Title: VA after cut items

The data source for this run = VA2.mdm
The command file for this run = C:\Users\NINJANAT\Desktop\Analysis VA after cut items\VA cut items model.hlm
Output file name = C:\Users\NINJANAT\Desktop\HLM value-added\hlm2.html
The maximum number of level-1 units = 2358
The maximum number of level-2 units = 223
The maximum number of iterations = 100

Method of estimation: restricted maximum likelihood

The outcome variable is SCIACH

Summary of the model specified

Level-1 Model

$$SCIACH_{mj} = \psi_{0j} + \psi_{1j}*(GENDER_{mj}) + \psi_{2j}*(EXTRA_{mj}) + \psi_{3j}*(ESCS_{mj}) + \psi_{4j}*(HEDRES_{mj}) + \psi_{5j}*(WEALTH_{mj}) + \psi_{6j}*(ICTRES_{mj}) + e_{mj}$$

Level-2 Model

$$\psi_{0j} = \gamma_{00} + \gamma_{01}*(PUBLIC_j) + \gamma_{02}*(TCSHORT_j) + \gamma_{03}*(FUNGOV_j) + \gamma_{04}*(SCMATEDU_j) + \gamma_{05}*(STRATIO_j) + \gamma_{06}*(RESPCURR_j) + u_{0j}$$

$$\psi_{1j} = \gamma_{10}$$

$$\psi_{2j} = \gamma_{20}$$

$$\psi_{3j} = \gamma_{30}$$

$\psi_{4j} = \gamma_{40}$
$\psi_{5j} = \gamma_{50}$
$\psi_{6j} = \gamma_{60}$

ESCS HEDRES WEALTH ICTRES have been centered around the grand mean.

TCSHORT FUNGOV SCMATEDU STRATIO RESPCURR have been centered around the grand mean.

Mixed Model

$$SCIACH_{mj} = \gamma_{00} + \gamma_{01}*PUBLIC_j + \gamma_{02}*TCSHORT_j + \gamma_{03}*FUNGOV_j + \gamma_{04}*SCMATEDU_j + \gamma_{05}*STRATIO_j + \gamma_{06}*RESPCURR_j + \gamma_{10}*GENDER_{mj} + \gamma_{20}*EXTRA_{mj} + \gamma_{30}*ESCS_{mj} + \gamma_{40}*HEDRES_{mj} + \gamma_{50}*WEALTH_{mj} + \gamma_{60}*ICTRES_{mj} + u_{0j} + e_{mj}$$

Final Results - Iteration 7

Iterations stopped due to small change in likelihood function

$\sigma^2_\varepsilon = 4533.91847$

τ

INTRCPT1,ψ_0	807.46755

Random level-1 coefficient	Reliability estimate
INTRCPT1,ψ_0	0.599

The value of the log-likelihood function at iteration 7 = -1.335561E+004

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *d.f.*	*p*-value
For INTRCPT1, ψ_0					
INTRCPT2, γ_{00}	469.839481	7.813026	60.135	216	<0.001
PUBLIC, γ_{01}	33.688074	7.888151	4.271	216	<0.001
TCSHORT, γ_{02}	3.669973	2.729984	1.344	216	0.180
FUNGOV, γ_{03}	-0.378655	0.114990	-3.293	216	0.001
SCMATEDU, γ_{04}	11.247036	3.021830	3.722	216	<0.001
STRATIO, γ_{05}	-0.331219	0.369984	-0.895	216	0.372
RESPCURR, γ_{06}	1.979199	3.172205	0.624	216	0.533

For GENDER slope, ψ_1					
INTRCPT2, γ_{10}	-2.110420	3.061793	-0.689	2129	0.491
For EXTRA slope, ψ_2					
INTRCPT2, γ_{20}	-4.104381	3.056957	-1.343	2129	0.180
For ESCS slope, ψ_3					
INTRCPT2, γ_{30}	13.024980	2.086607	6.242	2129	<0.001
For HEDRES slope, ψ_4					
INTRCPT2, γ_{40}	9.744525	2.314807	4.210	2129	<0.001
For WEALTH slope, ψ_5					
INTRCPT2, γ_{50}	-15.789469	2.784130	-5.671	2129	<0.001
For ICTRES slope, ψ_6					
INTRCPT2, γ_{60}	5.550783	2.019424	2.749	2129	0.006

Final estimation of fixed effects
(with robust standard errors)

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *d.f.*	*p*-value
For INTRCPT1, ψ_0					
INTRCPT2, γ_{00}	469.839481	8.117372	57.881	216	<0.001
PUBLIC, γ_{01}	33.688074	8.134529	4.141	216	<0.001
TCSHORT, γ_{02}	3.669973	2.632025	1.394	216	0.165
FUNGOV, γ_{03}	-0.378655	0.105666	-3.584	216	<0.001
SCMATEDU, γ_{04}	11.247036	3.132963	3.590	216	<0.001
STRATIO, γ_{05}	-0.331219	0.384361	-0.862	216	0.390
RESPCURR, γ_{06}	1.979199	3.008020	0.658	216	0.511
For GENDER slope, ψ_1					
INTRCPT2, γ_{10}	-2.110420	3.237258	-0.652	2129	0.515
For EXTRA slope, ψ_2					
INTRCPT2, γ_{20}	-4.104381	3.096316	-1.326	2129	0.185
For ESCS slope, ψ_3					
INTRCPT2, γ_{30}	13.024980	2.160025	6.030	2129	<0.001
For HEDRES slope, ψ_4					
INTRCPT2, γ_{40}	9.744525	2.493488	3.908	2129	<0.001
For WEALTH slope, ψ_5					
INTRCPT2, γ_{50}	-15.789469	2.557913	-6.173	2129	<0.001
For ICTRES slope, ψ_6					
INTRCPT2, γ_{60}	5.550783	2.110501	2.630	2129	0.009

Final estimation of variance components

Random Effect	Standard Deviation	Variance Component	*d.f.*	χ^2	*p*-value

INTRCPT1, u_0	28.41597	807.46755	216	653.07492	<0.001
level-1, e	67.33438	4533.91847			

Statistics for current covariance components model

Deviance = 26711.220597
Number of estimated parameters = 2

ภาคผนวก ง

คะแนนเฉลี่ยการประเมินการรู้เรื่องวิทยาศาสตร์ คะแนนมูลค่าเพิ่ม การจัดอันดับคะแนนมูลค่าเพิ่ม และการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่มระหว่าง 2 โมเดล

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางแสดง คะแนนเฉลี่ยการประเมินการรู้เรื่องวิทยาศาสตร์ คะแนนมูลค่าเพิ่ม การจัดอันดับคะแนนมูลค่าเพิ่ม และการจัดกลุ่มคุณภาพคะแนนมูลค่าเพิ่มระหว่าง 2 โมเดล

โรงเรียน	ก่อนพิจารณา DIF-DSF-DTF					หลังพิจารณา DIF-DSF-DTF				
	คะแนนเฉลี่ย	VAS	Rank	Rating	0/1	คะแนนเฉลี่ย	VAS	Rank	Rating	0/1
1	483.66	3.656899	81	4.0	1.0	490.84	5.149116	81	4.0	1.0
2	456.35	-11.1571	142	4.0	0.0	462.14	-10.9345	142	4.0	0.0
3	491.5	-14.5194	161	4.0	0.0	483.59	-14.4534	163	4.0	0.0
4	466.37	14.10966	59	3.0	1.0	472.49	15.99892	55	3.0	1.0
5	471.16	-22.5455	196	5.0	0.0	473.43	-24.3831	206	5.0	0.0
6	489.17	14.06524	60	3.0	1.0	487.46	11.36429	62	3.0	1.0
7	493.07	23.46796	44	3.0	1.0	490.64	26.51054	29	3.0	1.0
9	433.16	-5.03812	115	4.0	0.0	447.7	-5.44426	117	4.0	0.0
10	442.56	-37.7147	219	5.0	0.0	447.11	-34.5746	216	5.0	0.0
11	471.14	-30.7695	206	5.0	0.0	465.6	-32.0091	212	5.0	0.0
12	457.63	-16.7479	171	4.0	0.0	452.68	-22.1631	200	5.0	0.0
13	456.91	-11.6857	145	4.0	0.0	453.65	-16.4043	178	4.0	0.0
14	452.55	-23.6507	198	5.0	0.0	455.29	-21.5422	197	5.0	0.0
15	516.98	8.724157	69	3.0	1.0	512.44	8.519839	71	3.0	1.0
16	542.33	24.7779	39	3.0	1.0	535.35	19.78057	44	3.0	1.0
17	462.62	-17.1749	175	4.0	0.0	464.85	-21.4187	196	5.0	0.0
18	436.57	-23.0495	197	5.0	0.0	441.1	-19.526	191	4.0	0.0
19	552.52	31.16498	27	3.0	1.0	547.05	29.89479	23	3.0	1.0
20	551.56	34.86337	21	2.0	1.0	547.76	28.92688	25	3.0	1.0
21	466.29	-3.50566	105	4.0	0.0	458.37	-7.71028	126	4.0	0.0
22	460.58	-3.38058	104	4.0	0.0	456.86	-13.3645	157	4.0	0.0
24	494.47	0.20835	95	4.0	1.0	475.61	-12.2544	148	4.0	0.0
25	505.79	25.36675	37	3.0	1.0	504.18	24.90802	35	3.0	1.0
26	460.46	-32.1973	211	5.0	0.0	461.3	-32.086	213	5.0	0.0
27	423.75	-14.0188	158	4.0	0.0	420.23	-17.2218	180	4.0	0.0
28	468.64	-11.9202	146	4.0	0.0	467.72	-13.8823	159	4.0	0.0
29	538.23	11.22189	63	3.0	1.0	533.23	14.17958	59	3.0	1.0
30	442.5	-7.33821	126	4.0	0.0	448.98	-4.06523	106	4.0	0.0
31	530.07	-8.39303	128	4.0	0.0	535.06	-0.37141	96	4.0	0.0
32	528.72	41.41795	9	2.0	1.0	530.55	40.42601	10	2.0	1.0
33	499.38	-4.17859	108	4.0	0.0	503.36	-2.41579	103	4.0	0.0
34	473.7	-22.4188	195	5.0	0.0	478.79	-15.5454	171	4.0	0.0
35	434.45	-16.9494	173	4.0	0.0	439.55	-15.6515	173	4.0	0.0
36	433.96	-47.4002	222	5.0	0.0	436.89	-44.9958	222	5.0	0.0
37	465.02	27.08969	32	3.0	1.0	467.86	23.69003	37	3.0	1.0
38	483.03	-8.52349	131	4.0	0.0	478.45	-12.0197	146	4.0	0.0

โรงเรียน	ก่อนพิจารณา DIF-DSF-DTF					หลังพิจารณา DIF-DSF-DTF				
	คะแนนเฉลี่ย	VAS	Rank	Rating	0/1	คะแนนเฉลี่ย	VAS	Rank	Rating	0/1
39	522.14	-13.7235	156	4.0	0	521.39	-9.33559	136	4.0	0.0
40	414.97	-30.9151	207	5.0	0.0	428.25	-22.9615	203	5.0	0.0
41	407.87	3.190188	82	4.0	1.0	418.64	6.405104	79	3.0	1.0
42	531.62	20.61793	51	3.0	1.0	529.87	18.42079	47	3.0	1.0
43	607.4	39.61753	12	2.0	1.0	603.3	39.51003	11	2.0	1.0
44	457.13	-2.08416	100	4.0	0.0	460.39	-0.37793	97	4.0	0.0
45	448.02	-16.4305	170	4.0	0.0	451.87	-11.908	145	4.0	0.0
46	517.64	22.56176	45	3.0	1.0	499.13	12.7797	61	3.0	1.0
47	468.75	3.024415	83	4.0	1.0	483.63	5.716613	80	3.0	1.0
48	503.05	0.290657	93	4.0	1.0	487.29	0.448053	93	4.0	1.0
49	416.63	-17.217	176	4.0	0.0	415.86	-17.9942	182	4.0	0.0
50	438.82	-8.39553	129	4.0	0.0	437.85	-2.41446	102	4.0	0.0
51	439.99	-11.6162	144	4.0	0.0	437.78	-8.04464	129	4.0	0.0
52	559.22	21.83353	47	3.0	1.0	550.65	16.81292	52	3.0	1.0
53	484.1	4.01525	79	4.0	1.0	489.35	7.665737	75	3.0	1.0
54	483.55	17.03897	55	3.0	1.0	484.23	16.13267	54	3.0	1.0
55	473.81	16.58517	56	3.0	1.0	475.77	17.02435	51	3.0	1.0
56	439.9	-21.3207	192	5.0	0.0	440.93	-21.8444	199	5.0	0.0
57	460.42	5.735198	74	4.0	1.0	462.92	7.571039	76	3.0	1.0
58	448.76	-17.1135	174	4.0	0.0	457.67	-8.42009	132	4.0	0.0
59	503.05	-18.7999	182	4.0	0.0	497.64	-18.1487	184	4.0	0.0
60	445.68	-31.0921	208	5.0	0.0	451.24	-29.4962	210	5.0	0.0
61	532.72	21.55294	48	3.0	1.0	527.75	20.74395	41	3.0	1.0
62	463.33	5.470071	76	4.0	1.0	454.31	2.213932	87	4.0	1.0
63	490.96	-3.09795	103	4.0	0.0	495.79	-6.83167	123	4.0	0.0
64	424.6	-5.37044	117	4.0	0.0	422.43	-6.6208	122	4.0	0.0
65	604.02	25.49853	35	3.0	1.0	598.54	25.02355	33	3.0	1.0
66	426.68	-18.5131	180	4.0	0.0	423.29	-22.1702	201	5.0	0.0
67	592.71	28.97197	30	3.0	1.0	590.38	26.63604	28	3.0	1.0
68	481.56	12.8651	62	3.0	1.0	474.05	8.554803	69	3.0	1.0
69	458.29	-34.6129	215	5.0	0.0	463.57	-28.3629	208	5.0	0.0
70	445.04	-6.21566	120	4.0	0.0	443.38	-8.3355	131	4.0	0.0
71	479.08	-2.85563	102	4.0	0.0	475.06	-8.47522	133	4.0	0.0
72	581.76	38.01928	17	2.0	1.0	581.6	38.93892	12	2.0	1.0
73	463.95	-11.433	143	4.0	0.0	465.97	-10.4658	139	4.0	0.0
74	492.36	22.00645	46	3.0	1.0	486.12	13.9832	60	3.0	1.0
75	471.26	-20.2519	188	4.0	0.0	470.84	-19.0049	190	4.0	0.0
76	545.12	-4.37031	110	4.0	0.0	544.34	-1.80429	100	4.0	0.0
77	553.87	43.19808	6	2.0	1.0	556.83	47.3544	4	2.0	1.0
78	637.56	35.71651	20	2.0	1.0	620.68	27.44056	26	3.0	1.0

โรงเรียน	ก่อนพิจารณา DIF-DSF-DTF					หลังพิจารณา DIF-DSF-DTF				
	คะแนนเฉลี่ย	VAS	Rank	Rating	0/1	คะแนนเฉลี่ย	VAS	Rank	Rating	0/1
79	554.74	20.79557	50	3.0	1.0	570.88	26.00269	32	3.0	1.0
80	501.46	-15.7259	167	4.0	0.0	501.1	-12.1948	147	4.0	0.0
81	527.12	1.330798	89	4.0	1.0	527.29	3.753458	83	4.0	1.0
82	451.35	-6.12882	119	4.0	0.0	441.53	-11.1321	143	4.0	0.0
83	567.95	24.34516	40	3.0	1.0	565.3	23.05141	39	3.0	1.0
84	469.71	0.999614	92	4.0	1.0	470.76	2.689672	85	4.0	1.0
85	505.35	23.66134	43	3.0	1.0	500.13	16.1523	53	3.0	1.0
86	448.61	-33.993	214	5.0	0.0	412.57	-40.4087	221	5.0	0.0
87	442.9	-34.808	216	5.0	0.0	435.26	-38.708	219	5.0	0.0
88	527.85	20.94737	49	3.0	1.0	525.55	15.63023	57	3.0	1.0
89	489.83	-10.5653	140	4.0	0.0	491.11	-4.81887	112	4.0	0.0
90	416.92	-35.2702	218	5.0	0.0	419.41	-33.1297	214	5.0	0.0
91	456.21	-4.87316	113	4.0	0.0	444.27	-8.5026	134	4.0	0.0
92	468.96	-12.7267	153	4.0	0.0	471.3	-13.6296	158	4.0	0.0
93	449.2	-8.339	127	4.0	0.0	452.2	-5.83706	118	4.0	0.0
94	461.6	-1.89265	99	4.0	0.0	464.56	-0.54803	98	4.0	0.0
95	471.18	-27.3343	205	5.0	0.0	473.64	-24.3807	205	5.0	0.0
96	451.02	-4.6484	112	4.0	0.0	454.13	-7.58717	125	4.0	0.0
97	470.71	6.870105	73	3.0	1.0	481.04	10.45834	63	3.0	1.0
98	440.4	-25.4965	202	5.0	0.0	448.56	-18.3047	186	4.0	0.0
99	472.63	-13.7801	157	4.0	0.0	467.67	-18.1809	185	4.0	0.0
100	463.85	-12.7116	152	4.0	0.0	466.5	-5.20304	114	4.0	0.0
101	474.81	-6.29623	121	4.0	0.0	472.98	-10.3302	138	4.0	0.0
102	503.89	25.71912	34	3.0	1.0	500.34	19.8481	43	3.0	1.0
103	475.43	-2.76962	101	4.0	0.0	475.67	-3.47706	105	4.0	0.0
104	427.13	-32.1782	210	5.0	0.0	424.02	-28.5083	209	5.0	0.0
105	447.17	-14.7715	164	4.0	0.0	449.17	-14.7317	164	4.0	0.0
106	585.56	32.58431	24	3.0	1.0	579.51	28.96986	24	3.0	1.0
107	475.92	-10.3236	139	4.0	0.0	472.8	-15.4294	170	4.0	0.0
108	504.88	27.68837	31	3.0	1.0	492.1	17.94172	49	3.0	1.0
109	495.65	20.33364	52	3.0	1.0	499.56	19.35331	45	3.0	1.0
110	619.34	63.23851	2	1.0	1.0	619.55	62.96754	2	1.0	1.0
112	408.4	-1.7245	98	4.0	0.0	410.97	1.218632	90	4.0	1.0
113	465.49	13.61838	61	3.0	1.0	475.4	15.87482	56	3.0	1.0
114	407.98	-14.0808	160	4.0	0.0	410.94	-13.8915	160	4.0	0.0
115	446.12	-16.38	169	4.0	0.0	445.38	-22.5302	202	5.0	0.0
116	429.7	-14.0264	159	4.0	0.0	433.82	-14.909	166	4.0	0.0
117	534.94	15.53437	57	3.0	1.0	535.32	17.95221	48	3.0	1.0
118	437.8	-24.2397	200	5.0	0.0	440.84	-21.7709	198	5.0	0.0
119	533.32	36.93486	19	2.0	1.0	536.85	42.08884	8	2.0	1.0

โรงเรียน	ก่อนพิจารณา DIF-DSF-DTF					หลังพิจารณา DIF-DSF-DTF				
	คะแนนเฉลี่ย	VAS	Rank	Rating	0/1	คะแนนเฉลี่ย	VAS	Rank	Rating	0/1
120	509.45	1.576497	88	4.0	1.0	510.21	7.679199	74	3.0	1.0
121	503.2	39.90867	11	2.0	1.0	500.32	34.53914	17	2.0	1.0
122	473.14	2.52585	84	4.0	1.0	477.76	8.689006	68	3.0	1.0
123	454.51	1.932874	86	4.0	1.0	454.6	-1.72871	99	4.0	0.0
124	464.4	-20.3885	189	4.0	0.0	460.1	-18.8899	189	4.0	0.0
126	410.01	-3.84893	107	4.0	0.0	412.55	-6.47099	121	4.0	0.0
127	422.69	-8.75009	134	4.0	0.0	427.19	-6.02449	119	4.0	0.0
128	460.79	-12.3728	149	4.0	0.0	458.4	-12.9938	153	4.0	0.0
129	536.53	47.55761	5	2.0	1.0	535.65	46.64607	5	2.0	1.0
130	433.71	-4.32754	109	4.0	0.0	426.93	-8.09479	130	4.0	0.0
131	429.43	-18.0908	177	4.0	0.0	438.82	-15.9959	176	4.0	0.0
132	474.98	0.255176	94	4.0	1.0	472.34	2.601792	86	4.0	1.0
133	441.88	-5.61996	118	4.0	0.0	447.43	-4.86072	113	4.0	0.0
134	396.34	-31.7714	209	5.0	0.0	400.71	-28.1729	207	5.0	0.0
135	451.68	-14.6715	163	4.0	0.0	453.99	-15.0965	168	4.0	0.0
136	449.89	-20.1726	187	4.0	0.0	450.53	-13.2875	156	4.0	0.0
137	530.9	50.43958	3	2.0	1.0	527.14	48.8636	3	2.0	1.0
138	422.79	-23.8202	199	5.0	0.0	426.44	-19.8463	192	4.0	0.0
139	594.49	32.13068	25	3.0	1.0	587.57	27.26826	27	3.0	1.0
140	495.42	-5.06155	116	4.0	0.0	498.72	-6.20292	120	4.0	0.0
141	506.62	-19.574	186	4.0	0.0	503.26	-18.4341	187	4.0	0.0
142	592.39	38.10999	16	2.0	1.0	584.44	33.8444	20	2.0	1.0
143	552.8	49.5555	4	2.0	1.0	547.1	46.50959	6	2.0	1.0
144	417.97	-8.58489	132	4.0	0.0	427.28	-4.2228	108	4.0	0.0
145	473.18	1.04275	91	4.0	1.0	478.3	2.979976	84	4.0	1.0
146	476.18	10.46831	65	3.0	1.0	479.37	10.23685	64	3.0	1.0
147	499.15	2.335726	85	4.0	1.0	498.52	7.545941	77	3.0	1.0
148	441.46	-18.3211	178	4.0	0.0	445.2	-13.2096	155	4.0	0.0
150	441.53	-12.2886	147	4.0	0.0	445.39	-9.13464	135	4.0	0.0
151	525.14	41.75484	8	2.0	1.0	527.14	43.49911	7	2.0	1.0
152	539.75	38.39895	15	2.0	1.0	537.35	34.17505	19	2.0	1.0
153	500.98	-15.5185	166	4.0	0.0	499.83	-15.0313	167	4.0	0.0
154	512.08	-25.8125	203	5.0	0.0	506.26	-29.981	211	5.0	0.0
155	467.75	8.099363	71	3.0	1.0	466.63	0.181147	94	4.0	1.0
156	428.2	-35.0804	217	5.0	0.0	434.08	-24.2753	204	5.0	0.0
157	422.87	-19.1687	184	4.0	0.0	424.72	-19.9369	194	4.0	0.0
158	411.38	-14.5361	162	4.0	0.0	421.21	-10.8777	141	4.0	0.0
159	464.22	5.66757	75	4.0	1.0	469.22	8.340708	72	3.0	1.0
160	593.38	25.18306	38	3.0	1.0	587.51	23.82735	36	3.0	1.0
161	437.66	10.4599	66	3.0	1.0	437.84	6.876473	78	3.0	1.0

โรงเรียน	ก่อนพิจารณา DIF-DSF-DTF					หลังพิจารณา DIF-DSF-DTF				
	คะแนนเฉลี่ย	VAS	Rank	Rating	0/1	คะแนนเฉลี่ย	VAS	Rank	Rating	0/1
162	423.28	-24.6071	201	5.0	0.0	428.11	-16.5577	179	4.0	0.0
163	454.63	-12.3189	148	4.0	0.0	451.9	-13.0738	154	4.0	0.0
164	483.99	30.74158	28	3.0	1.0	482.57	26.15356	30	3.0	1.0
165	475.71	1.858313	87	4.0	1.0	478.23	1.845991	88	4.0	1.0
166	386.93	-26.2091	204	5.0	0.0	400.17	-20.1859	195	4.0	0.0
168	474.21	-12.5997	150	4.0	0.0	473.58	-12.6507	151	4.0	0.0
169	450.11	-10.0341	137	4.0	0.0	439.71	-15.3402	169	4.0	0.0
170	558.75	39.59224	13	2.0	1.0	557.02	38.45539	13	2.0	1.0
171	448.29	-14.8668	165	4.0	0.0	448.82	-10.732	140	4.0	0.0
172	498.84	29.56018	29	3.0	1.0	494.73	26.04607	31	3.0	1.0
173	504.05	42.74867	7	2.0	1.0	497.73	36.72822	16	2.0	1.0
174	526.96	23.89816	41	3.0	1.0	518.59	17.40328	50	3.0	1.0
175	467.72	-6.92794	124	4.0	0.0	469.48	-7.15346	124	4.0	0.0
176	441.29	-21.9141	193	5.0	0.0	448.14	-17.4496	181	4.0	0.0
177	417.37	-16.8503	172	4.0	0.0	423	-14.8954	165	4.0	0.0
178	572.15	23.76889	42	3.0	1.0	571.22	23.54062	38	3.0	1.0
179	413.57	-10.2344	138	4.0	0.0	429.62	-4.61049	110	4.0	0.0
180	496.73	-1.03548	96	4.0	0.0	498.09	1.54513	89	4.0	1.0
181	524.59	38.41141	14	2.0	1.0	530.72	41.76551	9	2.0	1.0
182	454.37	-3.68401	106	4.0	0.0	459.36	-2.79218	104	4.0	0.0
183	414.77	5.291501	77	4.0	1.0	429	8.549116	70	3.0	1.0
184	449.1	-18.5153	181	4.0	0.0	453.63	-15.9405	174	4.0	0.0
185	461.35	-5.02802	114	4.0	0.0	466.4	-4.20559	107	4.0	0.0
186	449.01	-6.68749	123	4.0	0.0	448.01	-5.30504	116	4.0	0.0
187	473.39	-8.67304	133	4.0	0.0	468.42	-9.70262	137	4.0	0.0
188	443.59	-7.30449	125	4.0	0.0	452.78	-4.26499	109	4.0	0.0
189	479.84	14.23669	58	3.0	1.0	485.88	15.41584	58	3.0	1.0
190	493.11	-18.9179	183	4.0	0.0	494.26	-16.1941	177	4.0	0.0
191	441.89	-9.84857	136	4.0	0.0	437.37	-5.2733	115	4.0	0.0
192	433.38	-47.6481	223	5.0	0.0	436.1	-46.0263	223	5.0	0.0
193	537.24	33.13185	23	3.0	1.0	534.33	30.92914	21	3.0	1.0
194	432.08	-22.246	194	5.0	0.0	435.73	-12.901	152	4.0	0.0
195	652.33	87.62351	1	1.0	1.0	644.68	82.42742	1	1.0	1.0
196	471.55	-13.365	154	4.0	0.0	466.4	-15.9567	175	4.0	0.0
197	445.33	5.191589	78	4.0	1.0	444.06	0.098414	95	4.0	1.0
198	475.6	10.67745	64	3.0	1.0	475.46	8.829607	67	3.0	1.0
199	439.12	-15.8782	168	4.0	0.0	438.93	-18.1204	183	4.0	0.0
200	448.95	-9.72848	135	4.0	0.0	456.88	-4.71582	111	4.0	0.0
201	416.9	-40.7213	220	5.0	0.0	420.3	-39.2457	220	5.0	0.0
202	561.42	37.31849	18	2.0	1.0	559.93	37.64576	15	2.0	1.0

โรงเรียน	ก่อนพิจารณา DIF-DSF-DTF					หลังพิจารณา DIF-DSF-DTF				
	คะแนนเฉลี่ย	VAS	Rank	Rating	0/1	คะแนนเฉลี่ย	VAS	Rank	Rating	0/1
203	545.66	-20.9414	191	5.0	0.0	543	-19.8573	193	4.0	0.0
205	447.48	-20.8285	190	5.0	0.0	456.63	-12.4597	149	4.0	0.0
206	476.74	17.80164	54	3.0	1.0	477.96	20.717	42	3.0	1.0
207	444.53	-33.955	213	5.0	0.0	443.64	-33.8464	215	5.0	0.0
208	546.99	-8.40849	130	4.0	0.0	546.3	-7.81071	127	4.0	0.0
209	458.16	-10.9536	141	4.0	0.0	454.85	-14.0564	161	4.0	0.0
210	506.83	33.24227	22	3.0	1.0	507.72	34.44138	18	2.0	1.0
211	469.37	26.55405	33	3.0	1.0	467.12	20.91781	40	3.0	1.0
212	433.6	7.55755	72	3.0	1.0	452.11	8.972334	66	3.0	1.0
213	461.9	-4.56951	111	4.0	0.0	457.23	-8.04192	128	4.0	0.0
214	393.83	-45.5285	221	5.0	0.0	404.38	-37.1521	218	5.0	0.0
215	519.91	25.42486	36	3.0	1.0	520.05	24.99701	34	3.0	1.0
216	474.24	8.220183	70	3.0	1.0	476.14	4.870843	82	4.0	1.0
217	554.97	40.51749	10	2.0	1.0	552.15	37.6963	14	2.0	1.0
218	473.83	1.117479	90	4.0	1.0	467.3	1.193623	91	4.0	1.0
219	417.4	-19.3466	185	4.0	0.0	430.23	-14.4346	162	4.0	0.0
220	441.01	18.93311	53	3.0	1.0	445.31	18.58748	46	3.0	1.0
221	441.48	9.070858	68	3.0	1.0	437.83	8.233305	73	3.0	1.0
222	507.27	-32.6499	212	5.0	0.0	498.17	-36.3035	217	5.0	0.0
223	500.29	-6.54575	122	4.0	0.0	496.25	-11.2458	144	4.0	0.0

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ประวัติผู้เขียนวิทยานิพนธ์

นายพนัส จันทร์เปล่ง เกิดที่บ้านโนนงิ้ว ต.คูเมือง อ.หนองบัวแดง จ.ชัยภูมิ สำเร็จการศึกษาปริญญาตรีวิทยาศาสตรบัณฑิต วิชาเอกคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ (ประสานมิตร) เมื่อปีการศึกษา 2545 และเข้าศึกษาต่อในหลักสูตรครุศาสตรมหาบัณฑิต สาขาการวัดและประเมินผลการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2546 และในปีการศึกษา 2553 ได้เข้าศึกษาศึกษาต่อในระดับปริญญาครุศาสตรดุษฎีบัณฑิต สาขาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ประวัติการทำงานในระหว่างศึกษาระดับปริญญาโทได้ทำหน้าที่เป็นผู้ช่วยสอน (TA) ประจำภาควิชาวิจัยการศึกษา(เดิม) คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย และเมื่อสำเร็จการศึกษา ระหว่างปี 2550 – 2551 เข้ารับตำแหน่งเป็นอาจารย์ สังกัดสำนักประกันคุณภาพ มหาวิทยาลัยเกษมบัณฑิต โดยมีภาระงานสอนในวิชาวิจัยธุรกิจและรับผิดชอบงานทางด้านการประกันคุณภาพ และในระหว่างปี 2552 – 2553 ได้ย้ายมารับตำแหน่งเป็นอาจารย์ สังกัดสำนักงานแผนและประกันคุณภาพการศึกษา มหาวิทยาลัยศรีปทุม โดยมีภาระงานสอนในวิชาคณิตศาสตร์เบื้องต้นและรับผิดชอบงานทางด้านการประกันคุณภาพร่วมด้วย

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University