การเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติภายใต้การหมุนแกน โครงสร้างเชิงมิติและระดับความสัมพันธ์ที่แตกต่างกัน

นางสาวพัชรี จันทร์เพ็ง

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาครุศาสตรดุษฎีบัณฑิต
สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา
คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2550

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

A COMPARISON OF QUALITY OF MULTIDIMENSIONAL ITEM RESPONSE THEORY LINKING METHODS UNDER THE DIFFERENCES OF ROTATION, DIMENSIONAL STRUCTURE, AND CORRELATION COEFFICIENT

Miss Putcharee Junpeng

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

A Dissertation Submitted in Partial Fulfillment of the Requirements

for the Degree of Doctor of Education Program in Educational Measurement and Evaluation

Department of Educational Research and Psychology

Faculty of Education Chulalongkorn University

Academic year 2007

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ ภายใต้การหมุนแกน โครงสร้างเชิงมิติ และระดับความสัมพันธ์ที่แตกต่างกัน
โดย	นางสาวพัชรี จันทร์เพ็ง
สาขาวิชา	การวัดและประเมินผลการศึกษา
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร. ณัฏฐภรณ์ หลาวทอง
อาจารย์ที่ปรึกษาร่วม	ศาสตราจารย์ ดร. ศิริชัย กาญจนวาสี

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาดุษฎีบัณฑิต

.....คณบดีคณะครุศาสตร์
(รองศาสตรจารย์ ดร. พฤทธิ์ ศิริบรรณพิทักษ์)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(รองศาสตรจารย์ ดร. ศิริเดช สุชีวะ)

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

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม
(ศาสตราจารย์ ดร. ศิริชัย กาญจนวาสี)

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

..... กรรมการ
(รองศาสตราจารย์ ดร.เอมอร จังศิริภรปกรณ์)

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

พัชรี จันทร์เพ็ง : การเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติภายใต้การหมุนแกน โครงสร้างเชิงมิติและระดับความสัมพันธ์ที่แตกต่างกัน. (A COMPARISON OF QUALITY OF MULTIDIMENSIONAL ITEM RESPONSE THEORY LINKING METHODS UNDER THE DIFFERENCES OF ROTATION, DIMENSIONAL STRUCTURE, AND CORRELATION COEFFICIENT) อ.ที่ปรึกษา : ผศ. ดร. ณัฏฐภรณ์ หลาวทอง, อ.ที่ปรึกษาร่วม : ศ. ดร. ศิริชัย กาญจนวาสี, 351 หน้า.

การวิจัยครั้งนี้เป็นการวิจัยเชิงทดลอง มีวัตถุประสงค์ 2 ประการ คือ 1) เพื่อศึกษาปฏิสัมพันธ์ของวิธีการเชื่อมโยงคะแนน โครงสร้างมิติความสามารถ และระดับความสัมพันธ์ระหว่างมิติความสามารถ และ 2) เพื่อเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนนตามโมเดล MIRT ที่มีลักษณะการหมุนแกนแตกต่างกัน 2 วิธี คือ การหมุนแกนแบบตั้งฉากโดยวิธี M (Min's Method) และการหมุนแกนแบบไม่ตั้งฉากโดยวิธี NOP (Non-Orthogonal Procrustes Method) ภายใต้โครงสร้างมิติความสามารถและระดับความสัมพันธ์ระหว่างมิติความสามารถที่ต่างกัน เงื่อนไขที่ทำการศึกษามี 12 เงื่อนไข (2X2X3) ประกอบด้วย วิธีการเชื่อมโยงคะแนน 2 วิธี คือ วิธี M และ วิธี NOP โครงสร้างมิติความสามารถ 2 ลักษณะ คือ โครงสร้างที่ไม่ซับซ้อน และโครงสร้างที่ซับซ้อนมากขึ้น และระดับความสัมพันธ์ระหว่างมิติความสามารถ 3 ระดับ คือ ไม่มีความสัมพันธ์ (r=0.0) มีความสัมพันธ์ระดับกลาง (r=0.5) และมีความสัมพันธ์ระดับสูง (r=0.8) ดำเนินการศึกษาจากการจำลองข้อมูลผู้สอบที่มีความสามารถต่างกัน 3 ระดับขั้น คือระดับต้น ระดับกลาง และระดับสูง ระดับขั้นละ 2,000 คน ออกแบบการเก็บข้อมูลสำหรับผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้ข้อสอบร่วมระดับละ 30 ข้อ ซึ่งศึกษาจากค่าพารามิเตอร์การแปลงคะแนนของข้อสอบจากระดับความสามารถต้นไปยังระดับกลาง และระดับกลางไปยังระดับสูง การจำลองข้อมูลมี 3 ขั้นตอน คือ 1) การสร้างค่าพารามิเตอร์ที่แท้จริงของข้อสอบและแบบแผนการตอบของนักเรียนแต่ละระดับ 2) การประมาณค่าพารามิเตอร์ตามโมเดล MIRT และ 3) การเชื่อมโยงคะแนนหรือการปรับเทียบคะแนน เกณฑ์ที่ใช้ในการเปรียบเทียบคุณภาพวิธีการเชื่อมโยงคะแนนพิจารณาจากดัชนีรากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) ดัชนีความลำเอียง (BIAS) และค่าความสัมพันธ์ระหว่างคะแนนที่แปลงได้กับคะแนนเกณฑ์ (CORR) พร้อมทั้งวิเคราะห์ความแปรปรวนแบบพหุจำแนก 3 ทาง (Three-way MANOVA) เพื่อเปรียบเทียบค่าเฉลี่ยของดัชนี RMSE BIAS และ CORR สำหรับศึกษาปฏิสัมพันธ์ระหว่างเงื่อนไขที่ศึกษา

สรุปผลการวิจัยได้ดังนี้

1. มีปฏิสัมพันธ์ระหว่างวิธีการเชื่อมโยงคะแนน (METHOD) โครงสร้างมิติความสามารถ (DS) และระดับความสัมพันธ์ระหว่างมิติความสามารถ (r) ซึ่งส่งผลต่อค่าความสัมพันธ์ระหว่างคะแนนที่แปลงได้กับคะแนนเกณฑ์(CORR) ของการแปลงค่าพารามิเตอร์จากระดับต้นไปยังระดับกลาง และส่งผลต่อความถูกต้องของการเชื่อมโยงคะแนน (BIAS) ของการแปลงค่าพารามิเตอร์จากระดับกลางไปยังระดับสูง อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่ไม่ส่งผลต่อความคงที่ของการเชื่อมโยงคะแนน (RMSE) ในทุกระดับของการแปลงคะแนน ส่วนปฏิสัมพันธ์รายคู่ พบว่า ปฏิสัมพันธ์ระหว่าง METHOD และ DS และปฏิสัมพันธ์ระหว่าง METHOD และ r ส่งผลต่อความคงที่และความถูกต้องของการเชื่อมโยงคะแนนที่ต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 กล่าวคือ คุณภาพของวิธีการเชื่อมโยงคะแนนที่ต่างกันขึ้นอยู่กับ DS และ r โดยมีปฏิสัมพันธ์ในทิศทางเดียวกันแต่ไม่ส่งผลให้มีค่าความสัมพันธ์ระหว่างคะแนนที่แปลงได้กับคะแนนเกณฑ์ (CORR) ต่างกัน

2. คุณภาพของการเชื่อมโยงคะแนนระหว่างวิธี M และวิธี NOP มีค่าใกล้เคียงกัน โดยเฉพาะเงื่อนไขที่ไม่มีความสัมพันธ์ระหว่างมิติความสามารถและมีความสัมพันธ์ระหว่างมิติความสามารถในระดับกลางกับโครงสร้างที่ไม่ซับซ้อน แต่เมื่อความสัมพันธ์ระหว่างมิติความสามารถเพิ่มขึ้นกับโครงสร้างที่ซับซ้อนมากขึ้น พบว่า วิธี NOP มีคุณภาพของการเชื่อมโยงคะแนนมากกว่าวิธี M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 นั่นคือ วิธี NOP มีความคงที่ (RMSE) และความถูกต้อง (BIAS) ของการเชื่อมโยงคะแนน มากกว่าวิธี M ซึ่งมีแนวโน้มในการประมาณค่าต่ำกว่าความเป็นจริง อย่างไรก็ตามผลการวิจัยพบว่าทั้งสองวิธีมีค่าความสัมพันธ์ระหว่างคะแนนที่แปลงได้กับคะแนนเกณฑ์ (CORR) ไม่ต่างกันไม่ว่าจะศึกษาในโครงสร้างลักษณะใดหรือความสัมพันธ์ระหว่างมิติความสามารถระดับใด

ภาควิชา	วิจัยและจิตวิทยาการศึกษา	ลายมือชื่อนิสิต.....
สาขาวิชา	การวัดและประเมินผลการศึกษา	ลายมือชื่ออาจารย์ที่ปรึกษา.....
ปีการศึกษา	2550	ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....

4884628927 : MAJOR EDUCATIONAL MEASUREMENT AND EVALUATION
KEY WORD: LINKING METHOD / MULTIDIMENSIONAL ITEM RESPONSE THEORY / MULTIDIMENSIONAL VERTICAL SCALING / DIMENSIONAL STRUCTURE

PUCHAREE JUNPENG : A COMPARISON OF QUALITY OF MULTIDIMENSIONAL ITEM RESPONSE THEORY LINKING METHODS UNDER THE DIFFERENCES OF ROTATION, DIMENSIONAL STRUCTURE, AND CORRELATION COEFFICIENT. THESIS ADVISOR : ASST.PROF. NATTAPORN LAWTHONG, Ph.D., THESIS COADVISOR : PROF. SIRICHAI KANJANAWASEE, Ph.D., 351 pp.

The purposes of this experimental study were to 1) compare the quality of the linking methods according to MIRT Model with two characteristics of rotations–the orthogonal rotation by M (Min's Method) and non-orthogonal rotation by NOP (Non-Orthogonal Procrustes Method)–under testing conditions of differences of dimensional structures and correlation levels between traits conditions 2) study the interaction of the linking methods, dimensional structures and correlation levels between traits.There were 12 conditions (2X2X3) consisted of two-linking methods–M and NOP methods, two-dimensional structures– approximate simple structure and mixed structure, three-correlation levels between traits conditions which were no correlation (r=0.0), moderate correlation (r=0.5) and high correlation (r=0.8). This research was based on the data simulation of the examinees with different ability classed in 3 levels; beginner, intermediate, and advance. There were 2,000 examinees in each level. The data collection for non-equivalent group design using the 20 common items in each level was studied by transforming items parameter from the beginner level to the intermediate level and from the intermediate level to the advance level. The data simulation consisted of three steps– generating true item parameters and response patterns for each grade level, calibrating MIRT parameter model and equating the procedure of linking. Root-mean squared error (RMSE), BIAS, and correlation coefficient between transformed estimated on the equated form and the criterion estimated on the base form (CORR) were used as criteria compare the quality of linking methods in this research. In addition, Three-way Multivariate Analysis of Variance (Three-way MANOVA) was conducted to compare means difference of RMSE, BIAS, and CORR in order to explore interaction of conditions being studied.

The results were as follows:

1. There was an interaction among linking method (METHOD), dimensional structure (DS), and correlation levels between traits (r) which affected correlation between transformed estimated and the criteria estimated (CORR) from the beginner level to the intermediate level and on accuracy of linking (BIAS) from the intermediate level to the advance level at the .05 significant level. However, the interaction had no effect on the stability of linking (RMSE) in every level of the items parameter transformation. The interaction between components showed that there were interactions between METHOD and DS, and METHOD and r which affected the difference between the stability and the accuracy of linking method at the .05 significant level. In other word, the differences of quality linking method depended on DS and r. The interactions between these conditions were ordinal interaction but they did not affect the difference of CORR.

2. The quality of linking methods between M and NOP, especially the conditions of no correlation and moderate correlation between traits which had approximate simple structure, was comparable. However, when the correlation became higher and the structure became much more complex, NOP had more quality of linking method than M method at .05 significant level. M method had more errors and under estimate parameters than NOP. Nevertheless, the result revealed that both correlation between transformed estimated and the criterion estimated (CORR) of M and NOP methods were not different in all conditions.

Department .....Educational Research and Psychology...Student's Signature.....Putcharee Junpeng.....
Field of Study .....Educational Measurement and Evaluation...Advisor's Signature ...Nattaporn Lawthong.....
Academic Year.....2007.....Co-advisor's Signature.....S. Kanjanawasee.....

กิตติกรรมประกาศ

การทำวิทยานิพนธ์ฉบับนี้ ผู้วิจัยได้รับความรู้และคำแนะนำต่างๆ ด้วยความกรุณาอย่างยิ่งจากอาจารย์ที่ปรึกษาวิทยานิพนธ์ คือ ผู้ช่วยศาสตรจารย์ ดร. ณัฏฐภรณ์ หลาวทอง และอาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม คือ ศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี ซึ่งอาจารย์ทั้งสองท่านได้ให้ความรู้ในหลักวิชามากมายอันทรงคุณค่ายิ่ง ทั้งยังได้คอยดูแลเอาใจใส่คอยติดตามความก้าวหน้าของผู้วิจัยตลอดมา ผู้วิจัยรู้สึกซาบซึ้งและขอกราบขอบพระคุณเป็นอย่างยิ่งด้วยความเคารพมา ณ ที่นี้ด้วย

ขอกราบขอบพระคุณ รองศาสตราจารย์ ดร. ศิริเดช สุชีวะ อาจารย์ ดร. ชูศักดิ์ ขัมภลิขิต และรองศาสตราจารย์ ดร. เอมอร จังศิริพรปกรณ์ ซึ่งเป็นกรรมการสอบวิทยานิพนธ์ที่ได้กรุณาสละเวลาอันมีค่ามาร่วมสอบวิทยานิพนธ์ ตลอดจนให้ข้อเสนอแนะที่เป็นประโยชน์ต่อการทำวิทยานิพนธ์ฉบับนี้จนสมบูรณ์ และขอกราบขอบพระคุณคณาจารย์ภาควิชาวิจัยและจิตวิทยาการศึกษา จุฬาลงกรณ์มหาวิทยาลัย ตลอดจนคณาจารย์สาขาวิชาการวัดและประเมินผลการศึกษา มหาวิทยาลัยขอนแก่น ที่คอยให้การสนับสนุนและเป็นกำลังใจด้วยดีตลอดมา

ขอขอบพระคุณว่าที่ร้อยโทภณัฐ ก้วยเจริญพานิชก์ ที่สละเวลาอันมีค่ายิ่ง ในการช่วยตรวจสอบความถูกต้องของโปรแกรมทางคอมพิวเตอร์ที่ผู้วิจัยพัฒนาขึ้นพร้อมทั้งให้คำแนะนำและข้อเสนอแนะที่เป็นประโยชน์ต่อการจำลองข้อมูลจนงานวิจัยเสร็จสมบูรณ์ และขอขอบพระคุณอาจารย์สังวรณ์ งัดกระโทก คุณชยุตม์ ภิรมย์สมบัติ Prof. Mark Reckase, Dr.Joseph Martineau, Dr.Haniza Yon Dr.Kyung-Seok Min และ Dr. Yuan Li ซึ่งเป็นผู้จุดประกายความคิดที่ทำให้ผู้วิจัยสนใจทำวิจัยเรื่องนี้พร้อมทั้งแนะนำเอกสารงานวิจัยและข้อเสนอแนะที่เป็นประโยชน์ด้วยดีเสนอมา และขอบคุณเป็นพิเศษสำหรับพี่ฝน เกรซ น้องเม และน้องส้มที่ตรวจสอบเรื่องสำนวนการเขียนวิทยานิพนธ์ของผู้วิจัย

งานวิจัยนี้จะไม่เสร็จสมบูรณ์ไปด้วยดี ถ้าไม่ได้รับความช่วยเหลือจาก พี่นงค์ พี่นุช พี่ต่าย และป้าน้อย โดยเฉพาะกำลังใจอันมีค่ายิ่งจากพี่ๆ ที่ร่วมศึกษาในระดับดุษฎีบัณฑิต อาธิ พี่สุนทร พี่มล พี่ต้อง พี่ดวง พี่ตุ๊ก พี่สมเกียรติ พี่หมี พี่ปอย พี่อู๋ พี่แอน พี่ภูมิ พี่เต็ม พี่สุรีพร พี่ผึ้ง พี่หมวย พี่ฝน พี่ปุ๊ก พี่หนู พี่แจง พี่อุ้ย พี่ปู พี่นาถ พี่เย่ และเพื่อนร่วมศึกษาในระดับมหาบัณฑิต อาธิ พี่น้อง พี่จุ่ง อ๊อด แจง พนัส ดิว จอม หนุ่ม รวมทั้งพี่ๆ และเพื่อนๆ ภาควิชาวิจัยและจิตวิทยาการศึกษาที่ให้กำลังด้วยดีตลอดมา

คุณค่าและประโยชน์ที่อาจมีต่อผู้อื่นของวิทยานิพนธ์ฉบับนี้ ผู้วิจัยขอมอบแก่ คุณพ่อธีระพล คุณแม่วิลัย คุณปู่ คุณย่า และคุณตา ตลอดจนญาติพี่น้องทุกคน ที่ได้ให้ความอบอุ่น ห่วงใยและสนับสนุนในทุกๆ ด้าน จนผู้วิจัยมีความสำเร็จมาถึงจุดนี้

สารบัญ

หน้า

บทคัดย่อภาษาไทย..... ง

บทคัดย่อภาษาอังกฤษ..... จ

กิตติกรรมประกาศ..... ฉ

สารบัญ..... ช

สารบัญตาราง..... ฌ

สารบัญภาพ..... ฏ

บทที่

1. บทนำ..... 1

ความเป็นมาและความสำคัญของปัญหา..... 1

คำถามการวิจัย..... 9

วัตถุประสงค์ของการวิจัย..... 10

สมติฐานการวิจัย..... 10

ข้อตกลงเบื้องต้น..... 12

ขอบเขตของการวิจัย..... 12

ข้อจำกัดของการวิจัย..... 13

คำกำจัดความที่ใช้ในการวิจัย..... 13

กรอบแนวคิดที่ใช้ในการวิจัย..... 19

ประโยชน์ที่คาดว่าจะได้รับ..... 20

2. เอกสารและงานวิจัยที่เกี่ยวข้อง..... 21

ตอนที่ 1 โมเดลการตอบสนองข้อสอบแบบพหุมิติ..... 21

ตอนที่ 2 แนวคิดเกี่ยวกับการเชื่อมโยงคะแนนตามโมเดลทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ และงานวิจัยที่เกี่ยวข้อง..... 49

ตอนที่ 3 การสร้างสเกลแนวตั้งพหุมิติ..... 113

3. วิธีดำเนินการวิจัย..... 134

ตอนที่ 1 เงื่อนไขที่ใช้ในการจำลองข้อมูล..... 136

ตอนที่ 2 การศึกษาการจำลองข้อมูล..... 141

ตอนที่ 3 การวิเคราะห์ข้อมูลที่ได้จากการจำลองข้อมูลและเกณฑ์ที่ใช้ในการเปรียบเทียบคุณภาพของการเชื่อมโยงคะแนน..... 154

ตอนที่ 4 การตรวจสอบความเป็นพหุมิติ และความถูกต้องของโปรแกรมที่ใช้ในการจำลองข้อมูล..... 155

หน้า

4. ผลการวิเคราะห์ข้อมูล..... 158
ตอนที่ 1 ผลการพิจารณาคุณภาพของการเชื่อมโยงคะแนนในภาพรวมโดยการวิเคราะห์ความแปรปรวนแบบพหุจำแนกตามวิธีการเชื่อมโยงคะแนน โครงสร้างมิติความสามารถและระดับความสัมพันธ์ระหว่างมิติความสามารถ.. 162
ตอนที่ 2 ผลการเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนนรายสถานการณ์... 202
ตอนที่ 3 สรุปผลการวิเคราะห์ตามวัตถุประสงค์การวิจัย..... 228
5. สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ..... 235
สรุปผลการวิจัย..... 236
อภิปรายผลการวิจัย..... 239
ข้อเสนอแนะ..... 246
รายการอ้างอิง..... 253
ภาคผนวก..... 265
ภาคผนวก ก รายชื่อผู้เชี่ยวชาญในการตรวจสอบโปรแกรมที่ผู้วิจัยพัฒนาขึ้น..... 266
ภาคผนวก ข ตัวอย่างคำสั่ง MATLAB สำหรับจำลองแบบแผนการตอบของผู้สอบ..... 268
ภาคผนวก ค ตัวอย่างแฟ้มข้อมูลแบบแผนการตอบของผู้สอบสำหรับกลุ่มผู้สอบขนาด 2,000 คน ข้อสอบร่วม 20 ข้อ..... 285
ภาคผนวก ง ตัวอย่างคำสั่ง MATLAB สำหรับคำนวณวิธีการเชื่อมโยงคะแนน..... 287
ภาคผนวก จ ตัวอย่างคำสั่งและผลการวิเคราะห์ค่าโอกาสการเดาของข้อสอบร่วม 30 ข้อ จากผู้สอบ 2,000 คน ด้วยโปรแกรม BILOG-MG..... 301
ภาคผนวก ฉ ตัวอย่างคำสั่งและผลการวิเคราะห์เพื่อประมาณค่าพารามิเตอร์อำนาจจำแนกและค่าความง่ายของข้อสอบ 30 ข้อ จากผู้สอบ 2,000 คนด้วยโปรแกรม TESTFACT..... 310
ภาคผนวก ช ผลการวิเคราะห์ค่าดัชนี RMSE BIAS และ สัมประสิทธิ์สหสัมพันธ์ในการทำซ้ำทั้งหมด 20 ครั้ง ในแต่ละเงื่อนไขที่ทำการศึกษา..... 338
ประวัติผู้เขียนวิทยานิพนธ์..... 351

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญตาราง

ตารางที่ หน้า

2.1 สรุปความแตกต่างของการวิเคราะห์องค์ประกอบ (Factor Analysis) และ MIRT..... 27

2.2 การจำแนกประเภทของการเชื่อมโยงคะแนนตามเกณฑ์ของ Feuer et al. (1999)..... 51

2. 3 ประเภทของการเชื่อมโยงคะแนนตามแนวคิดของ Mislevy (1992) & Linn (1993) และระดับของความคล้ายคลึงกันระหว่างแบบสอบสองฉบับ..... 58

2.4 ความแตกต่างของการประมาณค่าพารามิเตอร์ระหว่างโมเดล UIRT กับโมเดล MIRT..... 69

2.5 สรุปวิธีการและเงื่อนไขที่ใช้ในการเชื่อมโยงคะแนนตามโมเดล MIRT 90

2.6 สรุปงานวิจัยที่เกี่ยวข้องกับการเชื่อมโยงคะแนนตามโมเดล MIRT..... 108

3.1 จำนวนของข้อสอบต่อแบบสอบแต่ละชุดและขนาดของมุมในโครงสร้างAPSS... 139

3.2 จำนวนของข้อสอบต่อแบบสอบแต่ละชุดและขนาดของมุมในโครงสร้าง MS..... 140

3.3 ลักษณะการกระจายค่าความยากง่ายรวมและค่าอำนาจจำแนกรวมในแต่ละกลุ่มที่แตกต่างกันเพื่อใช้จำลองค่าพารามิเตอร์อำนาจจำแนก (a1 และ a2) และค่าความยากง่ายของข้อสอบ (d) สำหรับนักเรียนที่อยู่ในระดับที่สูงกว่า (Yon, 2006)..... 143

3.4 ลักษณะการกระจายค่าความยากง่ายรวมและค่าอำนาจจำแนกรวมในแต่ละกลุ่มที่แตกต่างกันเพื่อใช้จำลองค่าพารามิเตอร์อำนาจจำแนก (a1 และ a2) และค่าความยากง่ายของข้อสอบ (d) สำหรับนักเรียนที่อยู่ในระดับที่ต่ำกว่า (Yon, 2006)..... 143

3.5 ขนาดของมุมในมิติที่ 1 และมิติที่ 2 ซึ่งแยกตามโครงสร้างที่ไม่ซับซ้อน (APSS) และโครงสร้างที่มีความซับซ้อน (MS) เพื่อใช้สำหรับการสร้างค่าพารามิเตอร์อำนาจจำแนก (a1 และ a2) และค่าความยากง่ายของข้อสอบ (d) สำหรับการแปลงคะแนนในทุกระดับชั้น..... 145

3.6 ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ (a1 และ a2) และค่าความยากง่าย (d) ของข้อสอบร่วมจำนวน 30 ข้อ ของนักเรียนในระดับที่ 1 ทำแบบสอบในระดับที่ 2 และนักเรียนในระดับที่ 2 ทำแบบสอบในระดับที่ 2 ที่เหมาะสำหรับโครงสร้างที่ไม่ซับซ้อน (APSS) และโครงสร้างที่ซับซ้อนมากขึ้น (MS) เพื่อใช้สำหรับการแปลงคะแนนจากระดับที่ 1 (L1) ไปยังระดับที่ 2 (L2) 147

ตารางที่ หน้า

3.7 ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ (a1 และ a2) และค่าความยากง่าย (d) ของข้อสอบร่วมจำนวน 30 ข้อ ของนักเรียนในระดับที่ 2 ทำแบบสอบในระดับที่ 3 และนักเรียนในระดับที่ 3 ทำแบบสอบในระดับที่ 3 ที่เหมาะสำหรับโครงสร้างที่ไม่ซับซ้อน (APSS) และโครงสร้างที่ซับซ้อนมากขึ้น (MS) เพื่อใช้สำหรับการแปลงคะแนนจากระดับที่ 2 (L2) ไปยังระดับที่ 3 (L3)..... 148

3.8 การกระจายค่าความสามารถสำหรับผู้สอบแต่ละกลุ่ม..... 150

3.9 การออกแบบการเชื่อมโยงคะแนนสำหรับการจำลองข้อมูล: การใช้ข้อสอบจากกลุ่มผู้สอบที่ไม่เท่าเทียมกัน..... 151

4.1 ค่าสถิติเบื้องต้นสำหรับการพิจารณาดัชนี RMSE BIAS และค่าสัมประสิทธิ์สหสัมพันธ์ของผลที่ได้จากการแปลงค่าพารามิเตอร์ข้อสอบจากระดับที่ 1 ไปยังระดับที่ 2. 164

4.2 การทดสอบ Box's Test และ Bartlett's Test 171

4.3 การวิเคราะห์ความแปรปรวนพหุนามดัชนีบ่งชี้คุณภาพของการเชื่อมโยงคะแนนทั้ง 6 ตัวแปร ระหว่างวิธีการเชื่อมโยงคะแนน โครงสร้างมิติความสามารถและระดับความสัมพันธ์ระหว่างมิติความสามารถที่ต่างกัน..... 173

4.4 การทดสอบความแปรปรวนของค่าดัชนี RMSE BIAS และ CORR ของค่าอำนาจจำแนกในมิติที่ 1 และ มิติที่ 2 ทั้ง 6 ตัวแปร จาก Levene's Test 174

4.5 การทดสอบอิทธิพลระหว่างตัวแปรอิสระ (Test of Between-subject Effects) .. 178

4.6 ค่าสถิติเบื้องต้นสำหรับการพิจารณาดัชนี RMSE BIAS และ CORR ของผลที่ได้จากการแปลงค่าพารามิเตอร์ข้อสอบจากระดับที่ 2 ไปยังระดับที่ 3 183

4.7 การทดสอบ Box's Test และ Bartlett's Test..... 190

4.8 การวิเคราะห์ความแปรปรวนพหุนามดัชนีบ่งชี้คุณภาพของการเชื่อมโยงคะแนนทั้ง 6 ตัวแปร ระหว่างวิธีการเชื่อมโยงคะแนน โครงสร้างมิติความสามารถและระดับความสัมพันธ์ระหว่างมิติความสามารถที่ต่างกัน..... 191

4.9 การทดสอบความแปรปรวนของค่าดัชนี RMSE BIAS และ CORR ของค่าอำนาจจำแนกในมิติที่ 1 และ มิติที่ 2 ทั้ง 6 ตัวแปร จาก Levene's Test..... 192

4.10 การทดสอบอิทธิพลระหว่างตัวแปรอิสระ (Test of Between-subject Effects)... 194

4.11 สรุปผลการทดสอบอิทธิพลระหว่างตัวแปรอิสระ (Test of Between-subject Effects) ของการแปลงคะแนนจากระดับ 1 ไปยังระดับ 2 (L1_2) และ การแปลงคะแนนจากระดับ 2 ไปยังระดับ 3 (L2_3)..... 200

ตารางที่ หน้า

4.12 ผลการทดสอบความแปรปรปรวนแบบพหุจำแนกทางเดียว (1-WAY MANOVA) เพื่อเปรียบเทียบค่าเฉลี่ยของดัชนี Ln RMSE BIAS และค่า CORR ระหว่างวิธี M และวิธี NOP ในแต่ละเงื่อนไขที่ศึกษา จากระดับที่ 1 ไปยังระดับที่ 2 (L1_2)..... 203

4.13 ผลการทดสอบรายคู่ที่มีนัยสำคัญทางสถิติที่ระดับ .05..... 203

4.14 การเปรียบเทียบค่าเฉลี่ยของดัชนี RMSE BIAS และค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างวิธี M และวิธี NOP จากระดับที่ 1 ไปยังระดับที่ 2 (L1_2)..... 205

4.15 ผลการทดสอบความแปรปรวนแบบพหุจำแนกทางเดียว (1-WAY MANOVA) เพื่อเปรียบเทียบค่าเฉลี่ยของดัชนี Ln RMSE BIAS และค่า CORR ระหว่างวิธี M และวิธี NOP ในแต่ละเงื่อนไขที่ศึกษา จากระดับที่ 2 ไปยังระดับที่ 3 215

4.16 ผลการทดสอบรายคู่ที่มีนัยสำคัญทางสถิติที่ระดับ .05..... 215

4.17 การเปรียบเทียบค่าเฉลี่ยของดัชนี RMSE BIAS และค่า CORR ระหว่างวิธี M และวิธี NOP จากระดับที่ 2 ไปยังระดับที่ 3 (L2_3)..... 217

4.18 สรุปผลการทดสอบอิทธิพลระหว่างตัวแปรอิสระ (Test of Between-subject Effects) ของการแปลงคะแนนจากระดับ 1 ไปยังระดับ 2 (L1_2) และการแปลงคะแนนจากระดับ 2 ไปยังระดับ 3 (L2_3)..... 229

4.19 สรุปผลการทดสอบรายคู่ที่มีนัยสำคัญทางสถิติที่ระดับ .05..... 232

4.20 สรุปวิธีการเชื่อมโยงคะแนนที่มีค่าดัชนี RMSE น้อย BIAS ต่ำ และมีค่า CORR สูง ในภาพรวมทุกเงื่อนไข เมื่อเปรียบเทียบระหว่างวิธี M กับวิธี NOP..... 233

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญภาพ

แผนภาพที่ หน้า

1.1	กรอบแนวคิดในการวิจัย.....	19
2.1	Three-Dimensional Plot from the Multidimensional Logistic Model	33
2.2	การออกแบบกระบวนการทางโครงสร้างของ Wiener สำหรับ 4 เงื่อนไขของการวัด ตามโครงสร้างที่กำหนดไว้.....	37
2.3	เมตริกซ์ ของการออกแบบ MRMLC ซึ่งเป็นรูปแบบหนึ่งของโมเดล SLTM..	39
2.4	การแจกแจงคะแนนของแบบสอบที่มีลักษณะเป็นเอกมิติ	43
2.5	การแจกแจงคะแนนของแบบสอบที่มีลักษณะเป็นพหุมิติ	44
2.6	พื้นผิวการตอบข้อสอบ ซึ่งมี a1=1, a2=1, c=0.2 และ d=1.....	66
2.7	ระยะทางของมุมระหว่างมิติที่ 1 และมิติที่ 2.....	69
2.8	ส่วนประกอบของ UIRT Linking	73
2.9	ส่วนประกอบของ MIRT Linking (2 มิติ) จากการแปลงคะแนนองค์ประกอบ (Composite Transformation): A Rotation, a Translation, and a Central Dilation.....	74
2.10	Test Response Surfaces (TRs).....	77
2.11	แบบสอบ 2 มิติที่มีลักษณะโครงสร้างเชิงมิติอย่างง่าย.....	92
2.12	แบบสอบ 2 มิติที่มีลักษณะโครงสร้างที่ไม่ซับซ้อน.....	92
2.13	เวคเตอร์ข้อสอบ (Item Vectors) ที่มีลักษณะโครงสร้างที่ไม่ซับซ้อน.....	93
2.14	แบบสอบ 2 มิติที่มีลักษณะโครงสร้างที่ซับซ้อนมากขึ้น.....	94
2.15	เวคเตอร์ข้อสอบ (Item Vectors) ที่มีลักษณะโครงที่ซับซ้อนมากขึ้น.....	94
2.16	การออกแบบการเก็บรวบรวมข้อมูลโดยใช้รูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้แบบสอบร่วม.....	120
3.1	แบบแผนการจำลองข้อมูล.....	135
3.2	แบบสอบ 2 มิติที่มีลักษณะโครงสร้างที่ไม่ซับซ้อน.....	139
3.3	แบบสอบ 2 มิติที่มีลักษณะโครงสร้างที่ซับซ้อนมากขึ้น	140
3.4	ขั้นตอนการจำลองข้อมูลและการวิเคราะห์ข้อมูล.....	157
4.1	ค่าเฉลี่ยดัชนี RMSE BIAS และ CORR ของ a_1 ในการทำซ้ำ 20 ครั้ง ของ 6 เงื่อนไขที่ศึกษาเปรียบเทียบระหว่างวิธี M และ NOP	168
4.2	ค่าเฉลี่ยดัชนี RMSE BIAS และ CORR ของ a_2 ในการทำซ้ำ 20 ครั้ง ของ 6 เงื่อนไขที่ศึกษาเปรียบเทียบระหว่างวิธี M และ NOP	168
4.3	ค่าเฉลี่ยดัชนี RMSE BIAS และ CORR ของ d ในการทำซ้ำ 20 ครั้ง ของ 6 เงื่อนไขที่ศึกษาเปรียบเทียบระหว่างวิธี M และ NOP	169

แผนภาพที่ หน้า

4.4 กราฟแสดงปฏิสัมพันธ์ระหว่างวิธีการเชื่อมโยงคะแนน ระดับความสัมพันธ์ระหว่างมิติความสามารถในโครงสร้างที่ไม่ซับซ้อน (APSS) และโครงสร้างที่ซับซ้อนมากขึ้น (MS) ซึ่งมีผลต่อค่า CORR (METHOD X r at DS)…….....…….. 178

4.5 กราฟแสดงปฏิสัมพันธ์ METHOD X DS ในมิติที่ 1 และ 2……..... 179

4.6 กราฟแสดงปฏิสัมพันธ์ METHOD X r ……..... 180

4.7 ค่าเฉลี่ยดัชนี RMSE BIAS และ CORR ของ a_1 ในการทำซ้ำ 20 ครั้ง ของ 6 เงื่อนไขที่ศึกษาเปรียบเทียบระหว่างวิธี M และ NOP……. 186

4.8 ค่าเฉลี่ยดัชนี RMSE BIAS และ CORR ของ a_2 ในการทำซ้ำ 20 ครั้ง ของ 6 เงื่อนไขที่ศึกษาเปรียบเทียบระหว่างวิธี M และ NOP ……..... 187

4.9 ค่าเฉลี่ยดัชนี RMSE BIAS และ CORR ของ d ในการทำซ้ำ 20 ครั้ง ของ 6 เงื่อนไขที่ศึกษาเปรียบเทียบระหว่างวิธี M และ NOP ……... 188

4.10 กราฟแสดงปฏิสัมพันธ์ระหว่างวิธีการเชื่อมโยงคะแนน ระดับความสัมพันธ์ระหว่างมิติความสามารถในโครงสร้างที่ไม่ซับซ้อน (APSS) และโครงสร้างที่ซับซ้อนมากขึ้น (MS) ซึ่งมีผลต่อค่าดัชนี BIAS (METHOD X r at DS…….…….…… 196

4.11 กราฟแสดงปฏิสัมพันธ์ METHOD X DS……..…… 197

4.12 กราฟแสดงปฏิสัมพันธ์ METHOD X r ……..…… 198

4.13 แผนภูมิ Box Plot ของข้อมูลชุดวัดซ้ำทั้งหมด 20 ครั้ง สำหรับดัชนี RMSE ของค่าอำนาจจำแนกของข้อสอบในมิติที่ 1 (L1_2) ……..…… 206

4.14 แผนภูมิ Box Plot ของข้อมูลชุดวัดซ้ำทั้งหมด 20 ครั้ง สำหรับดัชนี RMSE ของค่าอำนาจจำแนกของข้อสอบในมิติที่ 2 (L1_2) …… 207

4.15 แผนภูมิ Box Plot ของข้อมูลชุดวัดซ้ำทั้งหมด 20 ครั้ง สำหรับดัชนี RMSE ของค่าความยากของข้อสอบ (L1_2) …… 208

4.16 แผนภูมิ Box Plot ของข้อมูลชุดวัดซ้ำทั้งหมด 20 ครั้ง สำหรับดัชนี BIAS ของค่าอำนาจจำแนกของข้อสอบในมิติที่ 1 (L1_2) …… 209

4.17 แผนภูมิ Box Plot ของข้อมูลชุดวัดซ้ำทั้งหมด 20 ครั้ง สำหรับดัชนี BIAS ของค่าอำนาจจำแนกของข้อสอบในมิติที่ 2 (L1_2) ……..…… 210

4.18 แผนภูมิ Box Plot ของข้อมูลชุดวัดซ้ำทั้งหมด 20 ครั้ง สำหรับดัชนี RMSE ของค่าความยากของข้อสอบ (L1_2)…….…… 211

4.19 แผนภูมิ Box Plot ของข้อมูลชุดวัดซ้ำทั้งหมด 20 ครั้ง สำหรับ CORR ของค่าอำนาจจำแนกของข้อสอบในมิติที่ 1 (L1_2)…… 212

แผนภาพที่ หน้า

4.20 แผนภูมิ Box Plot ของข้อมูลชุดวัดซ้ำทั้งหมด 20 ครั้ง สำหรับค่า CORR ของค่าอำนาจจำแนกของข้อสอบในมิติที่ 2 (L1_2)…… …..……..... 213

4.21 แผนภูมิ Box Plot ของข้อมูลชุดวัดซ้ำทั้งหมด 20 ครั้ง สำหรับดัชนี RMSE BIAS และ CORR สำหรับค่าพารามิเตอร์ d รายเงื่อนไขที่ศึกษา (L1_2).... 214

4.22 แผนภูมิ Box Plot ของข้อมูลชุดวัดซ้ำทั้งหมด 20 ครั้ง สำหรับดัชนี RMSE ของค่าอำนาจจำแนกของข้อสอบในมิติที่ 1 (L2_3)…….. 218

4.23 แผนภูมิ Box Plot ของข้อมูลชุดวัดซ้ำทั้งหมด 20 ครั้ง สำหรับดัชนี RMSE ของค่าอำนาจจำแนกของข้อสอบในมิติที่ 2 (L2_3)…….. 219

4.24 แผนภูมิ Box Plot ของข้อมูลชุดวัดซ้ำทั้งหมด 20 ครั้ง สำหรับดัชนี RMSE สำหรับค่าพารามิเตอร์ d รายเงื่อนไขที่ศึกษา (L2_3)……. 220

4.25 แผนภูมิ Box Plot ของข้อมูลชุดวัดซ้ำทั้งหมด 20 ครั้ง สำหรับดัชนี BIAS ของค่าอำนาจจำแนกข้อสอบในมิติที่ 1 (a1) รายเงื่อนไขที่ศึกษา (L2_3).... 221

4.26 แผนภูมิ Box Plot ของข้อมูลชุดวัดซ้ำทั้งหมด 20 ครั้ง สำหรับดัชนี BIAS ของค่าอำนาจจำแนกของข้อสอบในมิติที่ 2 (a2) รายเงื่อนไขที่ศึกษา …….. 222

4.27 แผนภูมิ Box Plot ของข้อมูลชุดวัดซ้ำทั้งหมด 20 ครั้ง สำหรับดัชนี BIAS สำหรับค่าพารามิเตอร์ d รายเงื่อนไขที่ศึกษา (L2_3)……. 223

4.28 แผนภูมิ Box Plot ของข้อมูลชุดวัดซ้ำทั้งหมด 20 ครั้ง สำหรับค่า CORR ของค่าอำนาจจำแนกของข้อสอบในมิติที่ 1 (a1) รายเงื่อนไขที่ศึกษา …….. 224

4.29 แผนภูมิ Box Plot ของข้อมูลชุดวัดซ้ำทั้งหมด 20 ครั้ง สำหรับค่า CORR ของค่าอำนาจจำแนกของข้อสอบในมิติที่ 2 (a2) รายเงื่อนไขที่ศึกษา……. 225

4.30 แผนภูมิ Box Plot ของข้อมูลชุดวัดซ้ำทั้งหมด 20 ครั้ง สำหรับค่า CORR ของพารามิเตอร์ d รายเงื่อนไขที่ศึกษา (L2_3)…….. 226

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1
บทนำ

ความเป็นมาและความสำคัญของปัญหา

การประเมินการเรียนรู้ตามแนวปฏิรูปการศึกษาเน้นการพัฒนาการเรียนรู้ของผู้เรียนให้ก้าวหน้าสูงสุด มีการปรับเปลี่ยน เพิ่มเติมเทคนิค วิธีการประเมินให้หลากหลายสอดคล้องกับเนื้อหาสาระ ถึงแม้ว่าจะให้ความสำคัญกับการประเมินด้วยเครื่องมือที่หลากหลาย แต่การประเมินด้วยแบบสอบยังเป็นวิธีการประเมินที่มีความจำเป็นในการวัดผลสัมฤทธิ์ทางการเรียน โดยเฉพาะความรู้และพัฒนาการความรู้ ซึ่งผลการวิเคราะห์ข้อสอบสามารถนำไปใช้วินิจฉัยปัญหาการเรียนรู้ได้ (เอมอร จังศิริพรปกรณ์, 2548) ซึ่งแบบสอบเป็นสิ่งเร้ารูปธรรมอย่างหนึ่งที่กระตุ้นให้ผู้สอบแสดงพฤติกรรมออกมาเป็นคำตอบ โดยนำคำตอบมาตีค่าเป็นคะแนน และคะแนนที่ตีค่าออกมานั้นไม่ได้บ่งบอกถึงความสามารถเดียวของผู้สอบ แต่อาจจะบ่งบอกอะไรได้อีกหลายๆ อย่าง ที่นอกเหนือจากสิ่งที่ผู้วัดต้องการวัดรวมอยู่ด้วย ดังนั้นการพัฒนาแบบสอบให้มีคุณภาพจึงเป็นสิ่งจำเป็นเพื่อให้การวัดมีค่าใกล้เคียงกับสภาพความเป็นจริงของพฤติกรรมหรือความสามารถของบุคคลในสิ่งที่ต้องการวัดมากที่สุด โดย Wright & Stone (1979) ได้กล่าวว่า คะแนนที่ได้จะมีความหมายนั้น จำเป็นต้องมาจากแบบแผนของการตอบสนองที่ประกอบด้วยข้อสอบที่ถูกกำหนดด้วยตัวแปรเดียว ดังนั้นการกำหนดแนวคิดของตัวแปรที่ต้องการวัดให้ชัดเจนนั้น ก็ควรระบุเพียงมิติเดียว ซึ่งในอดีตนักวัดผลให้ความสำคัญกับการพัฒนาแบบสอบให้มีลักษณะเป็นเอกมิติ ซึ่งมีความหลากหลายของรูปแบบการวัดที่แตกต่างกันแต่ต้องการวัดในสิ่งเดียวกัน โดยอาศัยเทคนิคของการเชื่อมโยงคะแนนระหว่างแบบสอบเพื่อทำให้เกิดความยุติธรรม เกิดความเที่ยงธรรม และให้ข้อมูลสารสนเทศที่เป็นประโยชน์เกี่ยวกับผู้สอบ ภายใต้ข้อตกลงเบื้องต้นของความเป็นเอกมิติ

ความจริงแล้วความเป็นเอกมิติของแบบสอบเกี่ยวกับเนื้อหาวิชา ที่ต้องการวัดนั้นออกจะกว้างมากจนแทบจะบอกความเป็นเอกมิติไม่ได้ และยิ่งในปัจจุบันนี้การเรียนการสอนในห้องเรียนที่เน้นถึงลักษณะของการบูรณาการความรู้มากขึ้น ทำให้การวัดมีวัตถุประสงค์เปลี่ยนแปลงไปจากอดีตโดยมุ่งเน้นการวัดสมรรถนะที่มีความซับซ้อนมากขึ้น โดยเฉพาะแบบสอบวัดผลสัมฤทธิ์ทางการเรียนในแต่ละกลุ่มสาระการเรียนรู้ เช่น กลุ่มสาระการเรียนรู้คณิตศาสตร์ที่มีกลุ่มเนื้อหาค่อนข้างเด่นชัด แบ่งออกเป็น 5 กลุ่ม คือ 1) จำนวนและการดำเนินการ 2) การวัด 3) เรขาคณิต 4) พีชคณิต และ 5) การวิเคราะห์ข้อมูลและความน่าจะเป็น แต่การเรียนรู้ในแต่ละกลุ่มของเนื้อหา จำเป็นต้องอาศัยการบูรณาการความรู้จากหลายกลุ่มเนื้อหา เช่น การเรียนรู้เกี่ยวกับพีชคณิตจำเป็นที่ผู้สอบต้องมีสมรรถนะทางด้านจำนวนและการดำเนินการ การเรียนรู้เกี่ยวกับเรขาคณิตจำเป็นที่ผู้สอบต้องมีสมรรถนะทางด้านการวัด เป็นต้น

เมื่อเป็นเช่นนี้ แนวปฏิบัติเกี่ยวกับการเรียนการสอนของครูจึงสวนทางกับแนวปฏิบัติของทฤษฎีทางด้านวัดผล (Torre & Patz, 2006) นิยามของเอกมิติก็คงต้องหลากหลายจนไม่อาจเป็นเอกมิติที่แท้จริงได้ โดย Warm (1978) ได้กล่าวว่า ข้อตกลงเบื้องต้นของเอกมิติมีความ

ซับซ้อนมากที่สุด และยังเป็นข้อตกลงที่เข้มงวดที่สุดของทฤษฎีการตอบสนองข้อสอบ (Item Response Theory; IRT) เมื่อเป็นเช่นนี้แล้ว ในฐานะที่เป็นนักวัดผลควรจะทำอย่างไร จึงได้เกิดมีข้อถกเถียงในเรื่องการวัดแบบสอบที่มีลักษณะพหุมิติ ซึ่งคะแนนที่ได้จะเป็นข้อมูลสำคัญเพื่อนำไปสู่การตัดสินใจทั้งระดับรายบุคคล ระดับสถาบัน รวมถึงระดับชาติ ในการนำไปเชื่อมโยงกับข้อมูลอื่น ซึ่งผลที่ได้ต้องมีความเสมอภาค และสามารถเปรียบเทียบกันได้ของคะแนน โดยถือว่าเป็นประเด็นสำคัญของสถานการณ์การทดสอบที่ได้รับความสนใจทั้งในอดีตและปัจจุบัน (Cook & Eignor, 1991) โดยเฉพาะความหลากหลายของรูปแบบการวัดที่แตกต่างกันแต่ต้องการวัดในสิ่งเดียวกัน (Kolen, 2001) อย่างไรก็ตามการเปรียบเทียบสเกลของแบบสอบเพื่อให้สามารถแปลความอย่างมีความหมาย เช่น ในกรณีที่มีระดับชั้นแตกต่างกัน ชนิดของแบบสอบแตกต่างกัน ในยุคที่ให้ความสำคัญกับการบูรณาการความรู้ จำเป็นอย่างยิ่งที่ต้องอาศัยเทคนิคของการเชื่อมโยงคะแนนระหว่างแบบสอบ (Test Linking Scores) โดยเฉพาะแบบสอบที่มีลักษณะเป็นแบบพหุมิติ

แนวคิดเกี่ยวกับ “**การเชื่อมโยงคะแนนระหว่างแบบสอบ**” (Test Linking Scores) มีอยู่หลายลักษณะ Mislevy (1992 อ้างใน Kolen & Brenan, 2004) และ Linn (1993 อ้างใน Kolen & Brenan, 2004) ได้จำแนกชนิดของการเชื่อมโยงคะแนนจากแบบสอบ 2 ฉบับ ใน “รูปแบบของการเชื่อมโยงคะแนน” โดยใช้ลักษณะร่วมกันของโครงสร้างเนื้อหาและความแกร่งจากผลของคะแนนที่ได้จากการเชื่อมโยงเป็นเครื่องมือในการจำแนก แบ่งออกเป็น 4 ชนิด คือ การปรับเทียบคะแนน (Equating), การปรับคะแนนให้เป็นมาตรฐาน (Calibration), การฉายภาพ (Projection) และการปรับค่าทางสถิติ (Statistical Moderation) ซึ่งกระบวนการเชื่อมโยงคะแนนจากแบบสอบ 2 ฉบับที่สำคัญ คือ การประเมินทั้งสองฉบับต้องมีเป้าหมายเหมือนกัน และมีเนื้อหาครอบคลุมลักษณะที่วัดเดียวกัน โดย Mislevy & Linn ได้กล่าวถึง **การปรับเทียบคะแนน (Equating)** ว่าเป็นการเชื่อมโยงคะแนนจากแบบสอบต่างฉบับ แต่แบบสอบถูกสร้างด้วยลักษณะของความร่วมกันในด้านโครงสร้างเนื้อหาและค่าสถิติระหว่างแบบสอบ 2 ฉบับ (Angoff, 1971) เมื่อพิจารณาเทียบกับ**การปรับคะแนนให้เป็นมาตรฐาน (Calibration)** หรือในทางปฏิบัตินิยมใช้คำว่า **การสร้างสเกลแนวตั้ง (Vertical Scaling)** เป็นการเชื่อมโยงคะแนนจากแบบสอบที่มีลักษณะร่วมกันของโครงสร้างเนื้อหา แต่มีระดับความยากที่แตกต่างกัน ถือได้ว่าเป็นการเชื่อมต่อที่มีเงื่อนไขน้อยกว่าการปรับเทียบคะแนน ส่วน**การฉายภาพ (Projection)** เป็นการเชื่อมโยงคะแนนจากแบบสอบที่มีโครงสร้างเนื้อหาต่างกันด้วยวิธีการทำนาย และ**การปรับด้วยค่าสถิติ (Statistical Moderation)** เป็นการเชื่อมโยงคะแนนจากแบบสอบที่มีโครงสร้างเนื้อหาต่างกัน มีตัวแปรปรับ (Moderator Variable) แฝงอยู่ในแบบสอบแต่ละฉบับ ซึ่งทั้งการฉายภาพและการปรับด้วยค่าสถิติเป็นการเชื่อมโยงคะแนนเมื่อข้อสอบวัดในโครงสร้างเนื้อหาต่างกัน จากการพิจารณาคำทั้ง 4 คำข้างต้น จะเห็นว่ามีความแตกต่างกันในข้อตกลงเบื้องต้นที่กำหนด โดยมีความแกร่งตามข้อตกลงเบื้องต้นลดน้อยลงตามลำดับ การปรับด้วยค่าสถิติมีข้อตกลงที่ผ่อนปรนมากที่สุด ส่วนข้อกำหนดของการปรับเทียบคะแนนจะมีข้อตกลงที่เข้มงวดมากที่สุด ซึ่งจากการศึกษา

เอกสารและงานวิจัยที่เกี่ยวข้อง พบว่าวิธีการปรับเทียบคะแนนระหว่างแบบสอบ (Test Equating) มีบทบาทสำคัญและเป็นเทคนิควิธีที่ได้รับการยอมรับตั้งแต่อดีตมาจนถึงปัจจุบัน

วิธีการปรับเทียบคะแนนได้มีผู้ศึกษาและพัฒนาอย่างต่อเนื่อง เมื่อพิจารณาจากการศึกษาที่ผ่านมา การปรับเทียบตามโมเดล IRT มีหลักฐานที่ให้ผลน่าเชื่อถือได้ (Slinde & Linn,1977-1979; Kolen, 1980; Kolen & Brennan, 2004; พรพิมล นาคเวช, 2537; พิชัย ละแมนชัย, 2549; วรรณ์ดี แสงประทีปทอง, 2538; ภัทรพร เกษสังข์, 2546; พัชรี จันทร์เพ็ง, 2547) เนื่องจากการปรับเทียบคะแนนตามโมเดล IRT ต้องมีความเสมอภาค (Equity) มีความสมมาตร (Symmetry) และไม่ผันแปรตามกลุ่ม (Invariance) นั่นคือ แบบสอบฉบับต่างๆ ที่นำมาปรับเทียบคะแนนกันต้องไม่ก่อให้เกิดผลที่แตกต่างกันกับผู้สอบ กล่าวโดยสรุปคะแนนที่ผ่านการปรับเทียบแล้วจะต้องมีการแจกแจงเหมือนกับคะแนนที่กลุ่มนั้นทำได้จริง ซึ่งคะแนนที่ได้ไม่ผันแปร ไม่ว่าจะปรับเทียบจากกลุ่มใด จึงสามารถนำผลไปใช้กับกลุ่มอื่นๆ ได้ และผลการปรับเทียบคะแนนจะต้องเหมือนกันไม่ว่าเป็นการปรับเทียบคะแนนจากแบบสอบฉบับที่ 1 ไปยังแบบสอบฉบับที่ 2 หรือปรับเทียบคะแนนจากแบบสอบฉบับที่ 2 ไปยังฉบับที่ 1 (ศิริชัย กาญจนวาสี, 2550)

อย่างไรก็ตามด้วยข้อจำกัดของความแกร่งต่อข้อตกลงเบื้องต้น (Strong Assumption) (Hambleton & Swaminathan, 1984; Embretson & Reise, 2000) และแบบสอบทั้งสองฉบับที่ทำการปรับเทียบกันต้องมุ่งวัดคุณลักษณะเดียวกัน ซึ่งเป็นการยากที่จะพบในสถานการณ์ของการสอบในปัจจุบันที่แบบสอบมีการวัดหลายมิติหรือหลายคุณลักษณะแฝง ส่งผลให้การรายงานผลของการปรับเทียบคะแนนในสเกลการวัดที่แตกต่างกันเกิดความไม่ถูกต้อง ด้วยปัญหาของการปรับเทียบคะแนนที่ต้องอยู่บนฐานความคิดดังกล่าว จึงมีการพัฒนาโมเดลการวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Item Response Theory; MIRT) เพื่อแก้ไขจุดบกพร่องของข้อตกลงเบื้องต้นดังกล่าว (Ackerman, 1992, 1994; Bock, Gibbons & Muraki, 1998; Reckase, 1985) ซึ่งโมเดลการวิเคราะห์ MIRT เป็นโมเดลที่มีความสอดคล้องกับข้อมูลมากกว่า ทำให้ผลการวิเคราะห์มีความถูกต้องและน่าเชื่อถือได้ ทั้งคุณสมบัติความไม่แปรเปลี่ยนของค่าพารามิเตอร์ความสามารถของผู้สอบและค่าพารามิเตอร์ของข้อสอบ นอกจากนี้จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง มีข้อสังเกตสำคัญประการหนึ่งคือ นักวิจัยทำการศึกษาวิธีการเชื่อมโยงคะแนน (Linking Score) ในรูปแบบของการปรับเทียบคะแนน แต่ไม่นิยมใช้ศัพท์คำว่าการปรับเทียบคะแนน (Equating) เหมือนเช่นในอดีตที่ผ่านมา แต่นิยมใช้คำว่าการเชื่อมโยงคะแนน (Linking) ดังนั้นในการวิจัยครั้งนี้เพื่อให้เกิดการยอมรับในระดับสากล ผู้วิจัยจึงใช้ศัพท์คำว่า *“การเชื่อมโยงคะแนนตามโมเดล MIRT (MIRT Linking)”* แทนคำว่า *“การปรับเทียบคะแนนตามโมเดล MIRT (MIRT Equating)”* แต่ยังเป็นการศึกษาในบริบทของการปรับเทียบคะแนน

วิธีการในการเชื่อมโยงคะแนนมีหลากหลายวิธี (Hirsch, 1989; Li & Lissitz, 2000; Thompson, Nering, & Davy, 1997; Oshima, Davy, & Lee, 2000; Min, 2003; Reckase & Martinue, 2004) ที่มีการพัฒนามาตามลำดับตั้งแต่อดีตถึงปัจจุบัน ซึ่งแต่ละวิธีนิยมออกแบบ

โดยใช้ข้อสอบร่วมเพื่อเป็นเครื่องมือในการเชื่อมโยงแบบสอบที่มีหลายฉบับให้สามารถอยู่บนสเกลที่สามารถเปรียบเทียบกันได้ อย่างไรก็ตามในแต่ละวิธีของการเชื่อมโยงคะแนนจะมีวิธีการที่เป็นเอกลักษณ์เฉพาะของคุณลักษณะของค่าสถิติ และมีเกณฑ์ที่มีความเหมาะสมมากที่สุดในแต่ละวิธี ซึ่งจากการสังเคราะห์วิธีการเชื่อมโยงคะแนนตามโมเดล MIRT สามารถแบ่งได้เป็น 2 ประเภท คือ วิธีการหมุนแกนแบบตั้งฉาก (Orthogonal Procrustes Rotation) กับวิธีการหมุนแกนแบบไม่ตั้งฉาก (Non-Orthogonal Procrustes Rotation) โดยมีรายละเอียด ดังนี้

1. การหมุนแกนแบบตั้งฉากมี 3 วิธี คือ 1) วิธีของ Hirsch (1989) ซึ่งถือเป็นบุคคลแรกที่ศึกษาวิธีการเชื่อมโยงคะแนนแบบพหุมิติในรูปแบบของการปรับเทียบคะแนน บนพื้นฐานของการออกแบบโดยใช้ผู้สอบร่วม แต่ถือเป็นวิธีการที่มีความยุ่งยากและซับซ้อนในทางปฏิบัติประกอบกับมีลักษณะในเชิงทดลอง 2) วิธี LL ซึ่ง Li & Lissitz (2000) ได้พัฒนากระบวนการเชื่อมโยงคะแนน 4 วิธีที่แตกต่างกัน ออกแบบโดยใช้ข้อสอบร่วม ผลที่ได้พบว่าวิธีการที่ดีที่สุดคือ การแปลงคะแนนรวม (Composite Transformation) ประกอบไปด้วยขั้นตอนการเชื่อมโยงคะแนนให้เกิดความยุติธรรม 3 ส่วนประกอบ คือ การหมุนแกน (Rotation) การแปลงคะแนน (Translation) และการเลื่อนแกนกลาง (Central Dilation) ถือได้ว่าเป็นส่วนประกอบที่มีความแกร่งในการให้ข้อมูลสารสนเทศเพื่อระบุขั้นตอนของการแปลงคะแนนแบบพหุมิติได้อย่างถูกต้อง รวมถึงมีการศึกษาค่าคงที่การเลื่อนแกนกลาง (Central Dilation) โดยมุ่งศึกษาค่าพารามิเตอร์การเลื่อนแกนที่เป็นสเกลลา (Scalar Dilation Parameter) และ 3) วิธี M ของ Min (2003) ที่ได้พัฒนาวิธีการที่เป็นรูปธรรมมากขึ้น จากการจำลองข้อมูลพร้อมทั้งนำไปประยุกต์ใช้กับข้อมูลจริงเพื่อศึกษาพัฒนาการผู้เรียน โดย Min พบจุดอ่อนที่สำคัญของวิธี LL (Li & Lissitz's Method) จากการพิจารณาค่าพารามิเตอร์การเลื่อนแกนที่มีลักษณะเป็นสเกลลา ว่าไม่เพียงพอในการหดหรือการขยายสเกลที่มีลักษณะหลากหลายมิติ ซึ่งค่าพารามิเตอร์การเลื่อนแกนที่มีลักษณะเป็นสเกลลาสามารถขยายสเกลไปยังมิติแต่ละมิติให้อยู่ในลักษณะที่เป็นรูปแบบเหมือนกันโดยตรง แต่เมื่อแยกการประมาณค่าในแต่ละมิติ อาจส่งผลให้สเกลมีการขยายไปยังมิติในระดับที่ต่างกัน มีปัญหาในการประมาณค่าสำหรับการปรับเทียบแนวตั้งที่มีหลายระดับที่แตกต่างกันได้ จึงทำการแทนค่าพารามิเตอร์การเลื่อนแกนที่มีลักษณะเป็นสเกลลาด้วยเมตริกซ์ (Diagonal Dilation Matrix) ทำให้ผลที่ได้มีประสิทธิภาพมากยิ่งขึ้น ซึ่งวิธี M แตกต่างจากวิธี LL (Li & Lissitz's Method) คือ ยอมให้มีการเปลี่ยนแปลงหน่วยความยาวในแต่ละมิติมากกว่าการเปลี่ยนแปลงค่าสเกลลาในมิติโดยรวมทั้งหมด ซึ่งค่าการเลื่อนแกนมีลักษณะเป็นเมตริกซ์ไม่ใช่สเกลลาเหมือนวิธี LL (Li & Lissitz's Method)

2. การหมุนแกนแบบไม่ตั้งฉากมี 2 วิธี คือ 1) วิธี OLD ซึ่ง Oshima Lee, & Davy (2000) ได้พัฒนาวิธีการเชื่อมโยงคะแนน (Oshima, Lee, & Davy Method; OLD) โดยใช้ข้อสอบร่วม ซึ่งชุดของข้อสอบร่วมจะรวมอยู่ในแบบสอบที่มีหลายฉบับเพื่อนำไปสู่สเกลร่วมกัน วิธีดังกล่าวไม่มีการประมาณค่าพารามิเตอร์การเลื่อนแกน (Dilation) แต่ศึกษาจากรูปแบบการแปลงคะแนนรวม (Composite Transformations) เพื่อนำมาใช้ในสถานการณ์เมตริกซ์การแปลง

คะแนนแบบไม่เป็นมุมฉาก (Nonorthogonal Transformation Matrix) และ 2) วิธี NOP (Non-Orthogonal Procrustes) ที่พัฒนาโดย Reckase & Martineau (2004) เป็นวิธีการแปลงคะแนนที่ต้องขึ้นอยู่กับความสอดคล้องมากที่สุดของมิติในแบบสอบปรับเทียบคะแนนไปยังแบบสอบฐาน ข้อได้เปรียบของวิธี NOP คือ มีการตัดเมตริกซ์การเลื่อนแกน (Dilation) ที่เป็นทั้งค่าเมตริกซ์หรือค่าสเกลลาออกจากสมการ ซึ่งวิธี NOP มีลักษณะคล้ายคลึงกับวิธี OLD (Oshima, Lee, & Davy's Method) ที่มิได้ทำการศึกษาเมตริกซ์การเลื่อนแกน เป็นการลดความยุ่งยากในระดับหนึ่ง โดยใช้กระบวนการทางคณิตศาสตร์มาช่วยในการคำนวณ พร้อมทั้งนำวิธีการไปประยุกต์ใช้เพื่อให้เกิดผลในทางปฏิบัติและมีประโยชน์จริงเพื่อการตัดสินใจเชิงนโยบาย โดยนำเทคนิคการเชื่อมโยงคะแนนตามโมเดล MIRT ในรูปแบบของการปรับเทียบคะแนน ไปประยุกต์ใช้ใน*การสร้างสเกลแนวตั้งแบบพหุมิติ* (Multidimensional Vertical Scaling) เพื่อศึกษาพัฒนาการของผลสัมฤทธิ์ทางการเรียนวิทยาศาสตร์ของนักเรียน 4 ระดับชั้นที่แตกต่างกันของมลรัฐมิชิแกน มีการออกแบบการเก็บรวบรวมข้อมูลรูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้แบบสอบร่วม

เมื่อพิจารณางานวิจัยที่เกี่ยวข้องตั้งแต่อดีตจนถึงปัจจุบัน (1989-2007) พบว่าเป็นช่วงเวลาสำคัญในการพัฒนาวิธีการเชื่อมโยงคะแนนที่เห็นผลในทางปฏิบัติและมีประโยชน์เพื่อประยุกต์ใช้ในสถานการณ์จริงเพิ่มขึ้น จึงเกิดประเด็นคำถามสำคัญโดยเฉพาะคำถามที่ว่า *"วิธีการใดที่มีความถูกต้องและเหมาะสมมากที่สุด?"* ซึ่งคำตอบที่ได้รับจากนักวัดผลและนักจิตมิติมักจะตอบคำถามในประเด็นนี้คือ *"ไม่มีวิธีการที่ดีที่สุด ต้องขึ้นอยู่กับเงื่อนไขของการปรับเทียบคะแนนในการนำไปเลือกใช้ในสถานการณ์ต่าง ๆ ที่แตกต่างกัน ในแต่ละบริบทที่ทำการศึกษา"* อย่างไรก็ตามคำตอบที่ได้ไม่สามารถตอบคำถามได้อย่างชัดเจนว่าวิธีการใดเหมาะสมกับสถานการณ์ใด โดยจากการศึกษาที่ผ่านมายังไม่มีการศึกษาเพื่อยืนยันฐานคิดทางทฤษฎีที่พัฒนาขึ้นอย่างจริงจัง โดยเฉพาะการศึกษาในเชิงเมตริกซ์การหมุนแกนที่ต่างกันระหว่างการหมุนแกนแบบตั้งฉาก และการหมุนแกนแบบไม่ตั้งฉาก ซึ่งเป็นองค์ประกอบที่สำคัญประการหนึ่งของการเชื่อมโยงคะแนนตามโมเดล MIRT โดยอาศัยระบบแกนอ้างอิง (Reference System) ดังที่ได้กล่าวมาแล้วข้างต้นว่า การหมุนแกนแบบตั้งฉากมี 3 วิธี ซึ่งวิธี M มีการพัฒนาเพื่อแก้ไขจุดอ่อนของ วิธี LL (Li & Lissitz's Method) ในขณะที่วิธีของ Hirsch (1989) มีความยุ่งยากและสลับซับซ้อนมากจนเกินไป ดังนั้น วิธี M (Min's Method) จึงมีความเหมาะสมในการนำมาศึกษาสำหรับกรณีของการหมุนแกนแบบตั้งฉากมากกว่าวิธีการอื่นที่มีการหมุนแกนในลักษณะเดียวกัน ในขณะที่การหมุนแกนแบบไม่ตั้งฉากมีหลายวิธีเช่นเดียวกัน แต่เนื่องจาก วิธี NOP (Non-Orthogonal Procrustes) เป็นวิธีที่มีความทันสมัยมากกว่าวิธีการเชื่อมโยงคะแนนแบบอื่น ๆ และพยายามแก้ไขจุดอ่อนของวิธีการในอดีต ดังนั้น วิธี NOP ถือได้ว่ามีความเหมาะสมมากกว่าวิธีการอื่นที่มีการหมุนแกนแบบไม่ตั้งฉาก ดังนั้นผู้วิจัยจึงสนใจเปรียบเทียบวิธี M และ วิธี NOP ซึ่งมีลักษณะการหมุนแกนที่ต่างกัน ประกอบกับทั้ง 2 วิธี ถือได้ว่าเป็นวิธีการที่พยายามแก้ไขจุดอ่อนและเพิ่มจุดเด่นของกระบวนการเชื่อมโยงคะแนน และมุ่งสู่การนำไปใช้ในทางปฏิบัติได้จริง

นอกจากนี้แบบสอบต่างๆ ทั้งทางด้านการศึกษาและจิตวิทยาโดยทั่วไปค่อนข้างเป็นการวัดที่ซับซ้อน และมีลักษณะเป็นองค์ประกอบเชิงซ้อน (Factorials Complex) ภายใต้การวัดหลายมิติ (McDonald, 1999) ในแต่ละมิติย่อมประกอบด้วยชุดของข้อสอบ ซึ่งข้อสอบข้อหนึ่งในบางสถานการณ์อาจสามารถวัดได้ในมิติหนึ่ง และเมื่อพิจารณาในอีกมุมหนึ่งสามารถวัดได้ในอีกมิติหนึ่งด้วยเช่นกัน สามารถพิจารณาในรูปของเวคเตอร์ข้อสอบ (Item Vector) ขึ้นอยู่กับว่าจะสามารถวัดในมิติใดมากกว่ากัน ตัวอย่างเช่น แบบสอบคณิตศาสตร์ชุดหนึ่งอาจจะประกอบไปด้วยการวัดพีชคณิต (Algebra) และเรขาคณิต (Geometry) ในกรณีนี้ชุดของแบบสอบย่อยมิติแรกกำหนดให้เป็นการวัดทางด้านพีชคณิต ส่วนมิติที่ 2 เป็นการวัดทางด้านเรขาคณิต โดยเป็นการศึกษาในลักษณะ 2 มิติ เมื่อพิจารณาในชุดของแบบสอบในแต่ละมิติ พบว่าความสัมพันธ์ระหว่างมิติสำหรับในสถานการณ์จริง มีความเป็นไปได้น้อยมากที่จะไม่มีความสัมพันธ์กันระหว่างมิติที่ทำการศึกษา โดย Zhang & Stout (1999) กล่าวว่าควรมีความจำเป็นอย่างยิ่งในการศึกษาโครงสร้างมิติความสามารถ (Dimensional Structure) ที่มีลักษณะหลากหลายมิติ โดยทั่วไปโครงสร้างมิติความสามารถประกอบไปด้วย 3 ลักษณะ คือ 1) โครงสร้างอย่างง่าย (Simple Structure; SS) มีลักษณะของชุดข้อสอบสามารถแบ่งแยกความแตกต่างของข้อสอบย่อยตามมิติใดมิติหนึ่งเพียงมิติเดียว อย่างไรก็ตามโครงสร้างดังกล่าวมักไม่สอดคล้องกับสภาพจริงเนื่องจากมีความเป็นไปได้น้อยมากที่ชุดข้อสอบจะขึ้นกับมิติใดมิติหนึ่งเพียงมิติเดียวเท่านั้น ดังนั้นโครงสร้างดังกล่าวจึงไม่เป็นที่นิยมในการนำมาศึกษา 2) โครงสร้างที่ไม่ซับซ้อน (Approximate Simple Structure; APSS) เป็นลักษณะโครงสร้างที่แต่ละข้อของแบบสอบมีความสัมพันธ์กันสูงในมิติที่หนึ่งมากกว่ามิติที่สอง และ 3) โครงสร้างที่ซับซ้อนมากขึ้น (Mixed Structure; MS) เป็นลักษณะโครงสร้างที่สอดคล้องกับสถานการณ์จริงมากกว่าโครงสร้างอย่างง่าย (SS) เนื่องจากโครงสร้างโดยทั่วไปไม่ได้มีเพียงมิติใดมิติหนึ่งเท่านั้น แต่ประกอบด้วยหลายๆ มิติ โดยลักษณะโครงสร้างที่ซับซ้อน (MS) จะมีการวัดในแต่ละมิติและมีการวัดมิติโดยรวม (Kim, 1994) ซึ่งมีลักษณะคล้ายกรณีของโครงสร้างที่ไม่ซับซ้อน (APSS) แต่ขนาดของมุมระหว่างข้อสอบกับแกนของมิติจะแตกต่างกันไป สถานการณ์โดยทั่วไปนิยมศึกษาโครงสร้างมิติความสามารถเพียง 2 ลักษณะ คือ โครงสร้างที่ไม่ซับซ้อน (APSS) และโครงสร้างที่ซับซ้อนมากขึ้น (MS) เนื่องจากสอดคล้องกับสถานการณ์จริงมากกว่าโครงสร้างอย่างง่าย (Roussos Stout & Marden, 1998; Min, 2003; Yon, 2006)

ประเด็นปัญหาสำคัญอีกประการหนึ่งที่ต่อเนื่องจากการวัดแบบสอบที่มีลักษณะแบบพหุมิติ ซึ่งเป็นข้อถกเถียงเกี่ยวการศึกษาในโมเดล MIRT เกี่ยวกับระดับความสัมพันธ์ระหว่างมิติความสามารถ (θ) ที่ทำการศึกษา โดยงานวิจัยส่วนใหญ่ที่ทำการศึกษา จะมีการกำหนดให้ไม่มีความสัมพันธ์ระหว่างมิติความสามารถ (r=0) นั่นคือระดับความสามารถของผู้สอบในอีกมิติหนึ่งไม่มีความสัมพันธ์กับความสามารถในอีกมิติหนึ่ง ซึ่งผู้วิจัยมีความเห็นว่าอาจไม่สอดคล้องกับสภาพความเป็นจริงในบางสถานการณ์ที่ทำการศึกษา ที่ระดับความสามารถของผู้สอบในอีกมิติหนึ่ง

มีความสัมพันธ์กับความสามารถในอีกมิติ จะมากหรือน้อยย่อมขึ้นอยู่กับเนื้อหาที่ทำการศึกษาเป็นสำคัญ โดยเฉพาะอย่างยิ่ง แบบสอบวัดผลสัมฤทธิ์ทางการเรียนที่มีการบูรณาการข้ามเนื้อหาสาระอย่างเห็นได้ชัด (Torre & Patz, 2006) ซึ่งจากการศึกษาของ CTB/McGraw-Hill (2002) และ Johnson & Carlson (1994) พบว่าแบบสอบวัดผลสัมฤทธิ์ทางการเรียนจะมีความสัมพันธ์ระหว่างมิติความสามารถอยู่ในระดับสูง มีค่าตั้งแต่ 0.7 ขึ้นไป โดยเมื่อนำมาหาค่าเฉลี่ยของความสัมพันธ์ระหว่างมิติความสามารถพบว่า อยู่ในระดับ 0.8 ซึ่งทำการศึกษาจากแบบสอบวัดผลสัมฤทธิ์ทางการเรียนเพื่อวัดความก้าวหน้าระดับชาติของประเทศสหรัฐอเมริกา (National Assessment of Education Progress; NAEP) ตั้งแต่ปี 1992-2007 เมื่อผู้วิจัยทำการวิเคราะห์กับหลักสูตรทางการศึกษาในประเทศไทยโดยเฉพาะในกลุ่มสาระการเรียนรู้คณิตศาสตร์ ที่มีการแบ่งเนื้อหาออกเป็น 5 กลุ่ม และมีการให้น้ำหนักความสำคัญในแต่ละกลุ่มเนื้อหาที่ใกล้เคียงกัน โดยมีความสัมพันธ์อยู่ในระดับสูงระหว่างมิติความสามารถเช่นเดียวกัน ดังนั้นผู้วิจัยจึงสนใจศึกษากับความสัมพันธ์ระหว่างมิติความสามารถในระดับสูงด้วย ซึ่งกำหนดให้มีค่าเท่ากับ 0.8 (r=0.8) เนื่องจากเป็นค่าที่มีความเป็นตัวแทนที่ได้จากการหาค่าเฉลี่ยของระดับความสัมพันธ์ที่มีค่าอยู่ในระดับสูง นอกจากนี้เพื่อให้การศึกษาในครั้งนี้มีความครอบคลุมมากขึ้นผู้วิจัยจึงสนใจศึกษากับความสัมพันธ์ระหว่างมิติความสามารถที่อยู่ในระดับกลางด้วย โดยกำหนดให้มีค่าเท่ากับ 0.5 (r=0.5) (Min, 2003)

นั่นคือ ในการศึกษาครั้งนี้มุ่งศึกษาวิธีการหมุนแกนที่ต่างกัน 2 วิธี คือ วิธีการหมุนแกนแบบตั้งฉากโดย วิธี M (Min's Method) และ วิธีการหมุนแกนแบบไม่ตั้งฉากโดยวิธี NOP (Non-Orthogonal Procrustes) โดยเงื่อนไขในการศึกษา คือ ระดับความสัมพันธ์ระหว่างมิติความสามารถ 3 ระดับ ประกอบด้วย ไม่มีความสัมพันธ์ (r=0.0) มีความสัมพันธ์ระดับกลาง (r=0.5) และมีความสัมพันธ์ระดับสูง (r=0.8) ที่มีโครงสร้างมิติความสามารถที่แตกต่างกัน 2 ลักษณะระหว่างโครงสร้างที่ไม่ซับซ้อนและโครงสร้างที่ซับซ้อนมากขึ้น เพื่อเป็นสารสนเทศสำคัญในการเลือกใช้วิธีการเชื่อมโยงคะแนนแบบพหุมิติตามโมเดล MIRT ในสถานการณ์ต่างๆ ได้อย่างถูกต้องและเหมาะสม เพื่อเป็นการขยายองค์ความรู้ และเป็นประโยชน์ต่อวงการศึกษาของประเทศไทยทั้งในปัจจุบันและอนาคต ที่มุ่งเน้นการประเมินความก้าวหน้าการเรียนรู้ของผู้เรียนที่มีการบูรณาการความรู้ข้ามเนื้อหา ซึ่งมีความจำเป็นต้องอาศัยทฤษฎีการวัดที่มีหลายมิติมาช่วยแก้ปัญหาดังกล่าว ดังนั้นผู้วิจัยจึงสนใจศึกษาวิธีการเชื่อมโยงแบบพหุมิติตามโมเดล MIRT ในรูปแบบของการปรับเทียบคะแนนไปประยุกต์ใช้ในบริบทของการสร้างสเกลแนวตั้ง (Multidimensional Vertical Scaling) ซึ่งเป็นวิธีการเชื่อมโยงคะแนนที่ได้รับความนิยมมากที่สุดในปัจจุบัน เพื่อนำมาใช้ในการศึกษาพัฒนาการความสามารถของผู้เรียนที่อยู่ในระดับชั้นต่างกันในบริบทของการวัดผลสัมฤทธิ์ทางการเรียนของผู้เรียน ซึ่งเป็นรูปแบบวิธีการเชื่อมโยงคะแนนที่มีความยืดหยุ่นและสอดคล้องกับสถานการณ์การทดสอบมากกว่าการปรับเทียบแนวตั้งแบบพหุมิติ (Multidimensional Vertical Equating) (Kolen & Brennan, 2004) นอกจากนี้พบว่าระบบการศึกษาในประเทศมีการแบ่งระดับการศึกษาออกเป็น 4 ช่วง

ชั้น ช่วงชั้นละ 3 ระดับชั้น ดังนั้นผู้วิจัยจึงสนใจศึกษาพัฒนาการทีละช่วงชั้น คือ ศึกษาใน 3 ระดับชั้นที่แตกต่างกัน โดยการศึกษาครั้งนี้สนใจศึกษาในสถานการณ์ของการจำลองข้อมูลมากกว่าการใช้ข้อมูลจริง เนื่องจากการศึกษาโดยจำลองข้อมูล ทำให้ทราบค่าพารามิเตอร์ที่แท้จริง ที่สามารถเปรียบเทียบค่าพารามิเตอร์ที่แท้จริงกับค่าที่ได้จากการแปลงคะแนนจากแบบสอบปรับเทียบคะแนนไปยังแบบสอบฐาน เพื่อเปรียบเทียบผลและความคลาดเคลื่อนที่เกิดขึ้นจากวิธีการเชื่อมโยงคะแนนที่ต่างกันในสถานการณ์ที่ทำการศึกษาได้ (Davey, Nering & Thompson, 1997; Bolt, 1999)

นอกจากนี้ผลการศึกษาที่ผ่านมาพบว่า กลุ่มผู้สอบประมาณ 2,000 คนขึ้นไปเหมาะสมสำหรับการศึกษาโดยโมเดล MIRT (Ackerman, 1994; Reckase, 1995; Yon, 2006) ดังนั้นในการศึกษาครั้งนี้จึงมุ่งเปรียบเทียบคุณภาพของการปรับเทียบคะแนนภายใต้เงื่อนไขที่ศึกษา โดยใช้กลุ่มตัวอย่างระดับชั้นละ 2,000 คน ออกแบบการเก็บรวบรวมข้อมูลรูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้ข้อสอบร่วม แม้ว่าจากการศึกษาที่ผ่านมายังไม่มีผลการศึกษาที่แน่ชัดว่าจำนวนข้อสอบร่วมที่เหมาะสมในการเชื่อมโยงคะแนนโดยโมเดล MIRT เป็นเท่าใด แต่จากการศึกษาโดยส่วนใหญ่นิยมใช้จำนวนข้อสอบประมาณ 20% ของข้อสอบทั้งหมด ตามการศึกษาจำนวนข้อสอบร่วมที่เหมาะสมของ Angoff (1971) & Budescu (1985) ซึ่งเป็นเกณฑ์ที่ได้รับการยอมรับในปัจจุบันเพื่อใช้ในการศึกษาโมเดล UIRT และ MIRT (อ้างใน Min, 2003; Yon, 2006) เนื่องจากผู้วิจัยได้ทำการศึกษาเบื้องต้น (Pilot Study) กับจำนวนข้อสอบร่วมเพียง 20 ข้อ พบว่ามีค่าความเที่ยงของข้อสอบต่ำโดยเฉพาะในเงื่อนไขที่ไม่มีความสัมพันธ์ระหว่างมิติความสามารถ (r=0.0) ซึ่งมีความเที่ยงอยู่ในช่วง 0.40-0.55 แต่เมื่อความสัมพันธ์ระหว่างมิติสูงขึ้นค่าความเที่ยงมีค่าเพิ่มขึ้นแต่มีค่าไม่เกิน 0.70 สาเหตุสำคัญคือข้อสอบมีน้อยเกินไปสำหรับใช้ในการศึกษา 2 มิติ ในกลุ่มตัวอย่างขนาดใหญ่ โดยผลการศึกษาดังกล่าวสอดคล้องกับการศึกษาของ Li & Lissize (2000) และ Reckase & Matineua (2000) ที่พบว่าจำนวนข้อสอบร่วมเป็นสิ่งที่สำคัญ โดยจำนวนข้อสอบร่วมที่ยาวกว่าย่อมให้ผลที่ดีกว่า ซึ่งจากการศึกษาของ Li & Lissitz (2000) ได้ศึกษากับจำนวนของข้อสอบร่วม 15 ข้อ และ 25 ข้อ ส่วน Reckase & Martineau (2004, 2006) ได้ใช้ข้อสอบร่วม ระดับละ 25 ข้อ ข้อค้นพบจากการศึกษาดังกล่าวพบว่าข้อสอบร่วมภายในน้อยเกินไปกับกลุ่มผู้สอบที่มีขนาดใหญ่ ดังนั้นการวิจัยในครั้งนี้ผู้วิจัยจึงเพิ่มจำนวนข้อสอบร่วมจาก 20 ข้อ เป็น 30 ข้อ ซึ่งคิดเป็น 30 % ของข้อสอบทั้งหมด

แม้ว่าในการศึกษาที่ผ่านยังไม่มีผลงานวิจัยที่ชัดเจนว่าควรมีการสร้างการจำลองข้อมูลโดยการทำซ้ำ (Replication) ทั้งหมดกี่ครั้ง เพื่อให้ผลที่เกิดขึ้นมีความเที่ยงมากที่สุด แต่จากการศึกษาของ Harwell, Hsu, & Kirisci (1996) พบว่าถ้าศึกษาโดยใช้โมเดล IRT เป็นฐานควรมีการทำซ้ำอย่างน้อย 20-25 ครั้ง และจากการศึกษางานวิจัยที่ทำการศึกษาโดยโมเดล MIRT จะมีการทำซ้ำประมาณ 10-20 รอบในแต่ละเงื่อนไขที่ศึกษา (Li & Lissitz, 2000; Min, 2003; Yon, 2006) ดังนั้นการศึกษาครั้งนี้จึงมีการสร้างค่าพารามิเตอร์ที่แท้จริงของข้อสอบและแบบแผนการตอบของนักเรียนแต่ละระดับโดยการทำซ้ำทั้งหมด 20 ครั้ง เพื่อให้การประมาณค่ามีความเที่ยงเพิ่มมากขึ้น

สำหรับเกณฑ์ที่ใช้ในการเปรียบเทียบคุณภาพครั้งนี้ ผู้วิจัยเลือกใช้ดัชนีรากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (Root-Mean Square Error; RMSE) และความลำเอียง (BIAS) ซึ่งเป็นดัชนีบ่งชี้คุณภาพของวิธีการเชื่อมโยงคะแนน เนื่องจากการใช้ดัชนีเหล่านี้เป็นเกณฑ์ในการเชื่อมโยงคะแนนสามารถแปลความหมายตามขนาดดัชนี ซึ่งเป็นค่าที่เรียงลำดับได้ว่าดัชนีที่มีค่าน้อยแสดงว่าความแตกต่างนั้นมีน้อย ซึ่งค่าดัชนีที่ได้มีลักษณะเป็นค่ามาตรฐาน เพราะถ่วงน้ำหนักด้วยความแปรปรวนของคะแนนเกณฑ์ ที่สามารถนำมาเปรียบเทียบกันได้โดยตรง ถึงแม้ในสถานการณ์ที่ได้ข้อมูลแตกต่างกันก็ตาม ซึ่งผลที่ได้จะเป็นสารสนเทศสำคัญในการเลือกใช้วิธีการเชื่อมโยงคะแนนที่เหมาะสมและมีความถูกต้อง นอกจากนี้มีการพิจารณาค่าความสัมพันธ์ระหว่างคะแนนที่แปลงได้จากแบบสอบปรับเทียบคะแนนและคะแนนเกณฑ์ที่ประมาณได้จากแบบสอบฐาน (CORR) พร้อมทั้งวิเคราะห์ความแปรปรวนแบบพหุ จำแนก 3 ทาง (Three-way MANOVA) เพื่อเปรียบเทียบค่าเฉลี่ยของ BIAS, RMSE และ CORR สำหรับศึกษาปฏิสัมพันธ์ของวิธีการเชื่อมโยงคะแนน โครงสร้างมิติความสามารถ และระดับความสัมพันธ์ระหว่างมิติความสามารถ ซึ่งผลที่ได้จากการวิจัยครั้งนี้จะเป็นประโยชน์เพื่อประกอบการตัดสินใจในการเลือกใช้วิธีการเชื่อมโยงคะแนนที่เหมาะสมและสอดคล้องกับสถานการณ์ที่ต้องการศึกษามากที่สุด ในการนำมาศึกษาพัฒนาการของผู้เรียนโดยศึกษาในแต่ละมิติเพื่อใช้ในการประเมินพัฒนาการผู้เรียนหรือประยุกต์ใช้ในการประเมินการเรียนการสอนของครู ตลอดจนประเมินผลสัมฤทธิ์ของโรงเรียนและหลักสูตร เพื่อเป็นสารสนเทศที่สำคัญทั้งในระดับสถานศึกษา เขตพื้นที่การศึกษา รวมถึงระดับประเทศ

คำถามการวิจัย

แม้ว่าแบบสอบทางด้านการศึกษาและจิตวิทยาโดยเฉพาะแบบสอบวัดผลสัมฤทธิ์ทางการเรียน จะมีลักษณะการวัดคุณลักษณะแฝงหลายมิติ ซึ่งจำเป็นต้องอาศัยโมเดล MIRT มาใช้ในการศึกษา ขณะที่วิธีการเชื่อมโยงคะแนนแบบพหุมิติตามโมเดล MIRT ได้ถูกพัฒนาขึ้นอย่างต่อเนื่องเพื่อนำมาใช้ในการแก้ปัญหาดังกล่าว (Hirsch, 1989; Oshima & Davy, 1994; Davy, Oshima, & Lee, 1996; Li, 1997; Li & Lissitz, 2000; Oshima, Davy, & Lee, 2000; Thompson, Nering, & Davy, 1997; Yon, 2006, 2007) ซึ่งงานวิจัยในอดีตยังไม่มีผลการวิจัยแน่ชัดว่าวิธีการใดเหมาะสมกับสถานการณ์การเชื่อมโยงคะแนนใด ประกอบกับจากการศึกษาของ Min (2003) ได้พัฒนาวิธีการเชื่อมโยงคะแนนที่มีลักษณะการหมุนแกนแบบตั้งฉาก (Orthogonal Procrustes Rotation) เรียกวิธีนี้ว่าวิธี M ในปีต่อมา Reckase & Martineau (2004) ได้พัฒนาวิธีการปรับเทียบคะแนนที่มีลักษณะการหมุนแกนแบบไม่ตั้งฉาก (Non-Orthogonal Procrustes Rotation) โดยเรียกวิธีนี้ว่า NOP ผู้วิจัยจึงเกิดข้อคำถามว่าวิธีการเชื่อมโยงคะแนนแบบพหุมิติ ที่มีลักษณะการหมุนแกนต่างกัน ระหว่างการหมุนแกนแบบตั้งฉากกับการหมุนแกนแบบไม่ตั้งฉากและมีโครงสร้างมิติความสามารถ (Dimensional Structure) แตกต่างกัน 2 ลักษณะ ระหว่างโครงสร้างที่ไม่ซับซ้อน (Approximate Simple Structure; APSS) และโครงสร้างที่ซับซ้อนมากขึ้น (Mixed Structure; MS) โดยแต่ละเงื่อนไขที่ทำการศึกษามีระดับความสัมพันธ์ของมิติความสามารถแตกต่างกัน 3 ระดับ

คือ ไม่มีความสัมพันธ์กัน (r=0) มีความสัมพันธ์กันในระดับกลาง (r=0.5) และมีความสัมพันธ์กันในระดับสูง (r=0.8) มีลักษณะเป็นอย่างไร โดยแยกเป็นประเด็นคำถามวิจัยไว้ดังนี้

1. วิธีการเชื่อมโยงคะแนน โครงสร้างมิติความสามารถและระดับความสัมพันธ์ระหว่างมิติความสามารถ มีปฏิสัมพันธ์ที่ส่งผลต่อคุณภาพของการเชื่อมโยงคะแนนหรือไม่อย่างไร เมื่อพิจารณาจากค่าเฉลี่ยของดัชนีค่าความลำเอียง (BIAS) ค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) และค่าความสัมพันธ์ระหว่างคะแนนที่แปลงได้กับคะแนนเกณฑ์ (CORR)

2. คุณภาพของการเชื่อมโยงคะแนนที่มีลักษณะการหมุนแกนแตกต่างกัน 2 วิธี คือ วิธีการหมุนแกนแบบตั้งฉากโดยวิธี M (Min's Method) และวิธีการหมุนแกนแบบไม่ตั้งฉากโดยวิธี NOP (Non-Orthogonal Procrustes Method) ภายใต้โครงสร้างมิติความสามารถและระดับความสัมพันธ์ระหว่างมิติความสามารถแตกต่างกัน จะมีคุณภาพแตกต่างกันหรือไม่ ผลการเชื่อมโยงคะแนนมีความสัมพันธ์กันหรือไม่ อย่างไร และเงื่อนไขใดเหมาะสมสำหรับการเชื่อมโยงคะแนนในสถานการณ์ใด โดยพิจารณาจากดัชนีค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) ค่าความลำเอียง (BIAS) และค่าความสัมพันธ์ระหว่างคะแนนที่แปลงได้จากแบบสอบปรับเทียบคะแนนกับคะแนนเกณฑ์ที่ประมาณได้จากแบบสอบฐาน (CORR)

วัตถุประสงค์การวิจัย

การวิจัยนี้มีวัตถุประสงค์หลักเพื่อเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนนแบบพหุมิติตามโมเดล MIRT เพื่อนำมาประยุกต์ใช้ในการสร้างสเกลแนวตั้งแบบพหุมิติ (Multidimensional Vertical Scaling) โดยมีจุดมุ่งหมายเฉพาะดังนี้

1. เพื่อศึกษาปฏิสัมพันธ์ของวิธีการเชื่อมโยงคะแนน และเงื่อนไขที่ทำการศึกษา คือ โครงสร้างมิติความสามารถ และระดับความสัมพันธ์ระหว่างมิติความสามารถที่ส่งผลต่อคุณภาพของการเชื่อมโยงคะแนน

2. เพื่อเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนนตามโมเดล MIRT ที่มีลักษณะการหมุนแกนแตกต่างกัน 2 วิธี คือ วิธีการหมุนแกนแบบตั้งฉากโดยวิธี M (Min's Method) และวิธีการหมุนแกนแบบไม่ตั้งฉากโดยวิธี NOP (Non-Orthogonal Procrustes Method) ภายใต้โครงสร้างมิติความสามารถและระดับความสัมพันธ์ระหว่างมิติความสามารถที่ต่างกัน

สมมติฐานการวิจัย

ในปี ค.ศ. 2003 Min ได้ทำการศึกษาปฏิสัมพันธ์ระหว่างวิธีการเชื่อมโยงคะแนนกับเงื่อนไขของการศึกษาคือ ขนาดกลุ่มตัวอย่าง, ค่าการกระจายความสามารถของกลุ่มตัวอย่าง และโครงสร้างมิติความสามารถ จากการวิเคราะห์โดยใช้ Repeated Measure ANOVA จากค่าความคลาดเคลื่อนของวิธีการเชื่อมโยงคะแนนในแต่ละรอบ ที่พิจารณาจากค่าเฉลี่ยของดัชนี BIAS RMSE และค่าสัมประสิทธิ์สหสัมพันธ์ ซึ่งถือเป็นคุณภาพของวิธีการเชื่อมโยงคะแนน

ผลการวิจัยพบว่า อิทธิพลของวิธีการเชื่อมโยงคะแนนขึ้นอยู่กับเงื่อนไขที่ทำการศึกษา นั่นคือมีปฏิสัมพันธ์อย่างมีนัยสำคัญทั้งระหว่างองค์ประกอบและภายในองค์ประกอบที่ทำการศึกษา (Between-and Within-Factor) โดยองค์ประกอบหลักคือวิธีการเชื่อมโยงคะแนน และเงื่อนไขที่ทำการศึกษามีปฏิสัมพันธ์อย่างมีนัยสำคัญต่อคุณภาพของวิธีการเชื่อมโยงคะแนน นอกจากนี้ในปี ค.ศ. 2006 Yon ได้ทำการศึกษาปฏิสัมพันธ์ระหว่างวิธีการเชื่อมโยงกับโครงสร้างมิติความสามารถจากการวิเคราะห์โดย Repeated Measure ANOVA ที่พิจารณาค่าเช่นเดียวกับการศึกษาของ Min (2003) ซึ่งให้ผลที่มีความสอดคล้องกัน จากข้อค้นพบที่ได้จึงนำไปสู่การตั้งสมมติฐานเพื่อตอบคำถามวิจัยข้อที่ 1 ดังต่อไปนี้

1. วิธีการเชื่อมโยงคะแนนและเงื่อนไขที่ทำการศึกษาคือ โครงสร้างมิติความสามารถและระดับความสัมพันธ์ของมิติความสามารถ น่าจะมีปฏิสัมพันธ์ต่อกันซึ่งส่งผลต่อคุณภาพของการเชื่อมโยงคะแนน อย่างมีนัยสำคัญทางสถิติ เมื่อพิจารณาจากค่าเฉลี่ยของดัชนีค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) ค่าความลำเอียง (BIAS) และค่าความสัมพันธ์ระหว่างคะแนนที่แปลงได้กับคะแนนเกณฑ์ (CORR)

นอกจากนี้ ผลการวิจัยของ Min (2003) พบว่า วิธี M เหมาะสมสำหรับมิติที่ทำการศึกษามีจำนวน 2-3 มิติ (Low-Dimensional Space) แต่สำหรับกรณีที่โมเดลมีจำนวนมิติมากขึ้นและมีความซับซ้อนมากเกินไป (High-Dimensional Space) การคำนวณจัดกระทำได้ยากหรือไม่สามารถคำนวณได้ เนื่องจากปฏิสัมพันธ์ระหว่างคุณลักษณะเด่นของมิติข้ามชุดของแบบสอบโดยเฉพาะอย่างยิ่งคุณลักษณะของการหมุนแกนแบบตั้งฉาก ดังนั้นจำเป็นที่จะต้องมีการหมุนแกนการเปลี่ยนแปลงมิติของการเปรียบเทียบแต่ละคู่จากแบบสอบปรับเทียบคะแนนไปยังแบบสอบฐาน ดังนั้นวิธี M จึงน่าจะเหมาะสมกับโครงสร้างมิติความสามารถที่ไม่ซับซ้อน ในขณะที่วิธี NOP (Reckase & Martineau, 2004) ซึ่งมีการหมุนแกนแบบตั้งฉาก โดยมีการหมุนแกนเป็นไปโดยอัตโนมัติสำหรับการแปลงคะแนนในแต่ละมิติของเมตริกซ์ปรับเทียบคะแนนกับมิติของเมตริกซ์ฐาน ขึ้นอยู่กับความสอดคล้องของมิติในเมตริกซ์ปรับเทียบคะแนนกับมิติของเมตริกซ์ฐาน ขณะเดียวกันการแปลงคะแนนแบบไม่ตั้งฉาก ช่วยลดการคำนวณน้ำหนักของการปรับเทียบคะแนนทุกครั้งของการแปลงคะแนน ในเมตริกซ์ปรับเทียบคะแนนช่วยลดความยุ่งยากในระดับหนึ่ง ซึ่งการตัดสินใจของมิติในเมตริกซ์ปรับเทียบคะแนน คือการจับคู่กับมิติในเมตริกซ์ฐาน โดยใช้กระบวนการทางคณิตศาสตร์มาช่วยคำนวณ ข้อดี ของการแปลงคะแนนแบบไม่ตั้งฉาก คือการช่วยลดความจำเป็นหรือลดบทบาทของค่าพารามิเตอร์การเลื่อนแกน (Dilation) โดยเหมาะสมกับแบบสอบที่มีจำนวนมิติมากขึ้น วิธี NOP จึงน่าจะเหมาะสมกับโครงสร้างมิติความสามารถที่มีความซับซ้อนมากกว่าวิธี M

เมื่อพิจารณาระดับความสัมพันธ์ของมิติความสามารถ พบว่า วิธี M มีลักษณะการหมุนแกนโดยทำมุมฉากระหว่างมิติ ดังนั้นจึงน่าจะเหมาะกับมิติความสามารถที่ไม่มีความสัมพันธ์ระหว่างกัน ในขณะที่วิธี NOP ที่มีลักษณะการหมุนแกนที่ทำมุมแหลมระหว่างระหว่างมิติ ที่อาจเกิด

ความสัมพันธ์ระหว่างมิติความสามารถที่ทำการศึกษา ดังนั้น จึงน่าจะเหมาะกับระดับความสัมพันธ์ที่มีค่าในระดับกลางหรือสูงขึ้นไป จากข้อค้นพบที่ได้จึงนำไปสู่การตั้งสมมติฐานเพื่อตอบคำถามวิจัยข้อที่ 2 ดังต่อไปนี้

2. คุณภาพการเชื่อมโยงคะแนนตามวิธีการหมุนแกนแบบตั้งฉากโดยวิธี M (Min's Method) น่าจะเหมาะสมสำหรับโครงสร้างที่ไม่ซับซ้อน (Approximate Simple Structure; APSS) และมิติความสามารถที่ไม่มีความสัมพันธ์ระหว่างกัน (r=0) ในขณะที่วิธีการหมุนแกนแบบไม่ตั้งฉากโดยวิธี NOP (Non-Orthogonal Procrustes Method) น่าจะเหมาะสมกับโครงสร้างที่ซับซ้อนมากขึ้น (Mixed Structure; MS) กับระดับความสัมพันธ์ที่มีค่าในระดับกลาง (r=0.5) หรือระดับสูง (r=0.8) โดยพิจารณาจากดัชนีค่าความลำเอียง (BIAS) ค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) และค่าความสัมพันธ์ระหว่างคะแนนที่แปลงได้กับคะแนนเกณฑ์ (CORR)

ข้อตกลงเบื้องต้น

1. สถานการณ์การจำลองข้อมูลในครั้งนี้ มีข้อตกลงเบื้องต้นว่ากลุ่มตัวอย่างได้มาจากการสุ่มของประชากร และระดับความสามารถของคุณลักษณะแฝงของผู้สอบมาจากการแจกแจงแบบโค้งปกติที่มีลักษณะทวิปรกติ (Bivariate Normal Distribution) นั่นคือระดับความสามารถมีค่าเฉลี่ยเป็น 0 และมีส่วนเบี่ยงเบนมาตรฐานเป็น 1 ซึ่งความสามารถที่แท้จริงจะถูกสร้างขึ้นจากกลุ่มตัวอย่างของนักเรียนในแต่ละระดับชั้นจากข้อมูลที่มีการแจกแจงปกติ หรือ $N(\mu_g,\sigma_g)$ เมื่อ g คือระดับชั้นที่ศึกษา โดยในการศึกษาครั้งนี้เป็นระดับชั้นที่ 1, 2 และ 3

2. การวิจัยครั้งนี้กำหนดให้ชุดของข้อสอบมีการกระจายค่าความยากโดยรวม (MDISC) และค่าอำนาจจำแนกโดยรวม (MDIFF) ในแต่ละกลุ่มที่แตกต่างกัน 5 ระดับ ในแต่ละระดับจะมีค่าความยากที่แตกต่างกัน ซึ่งมีการกำหนดค่าที่เป็นไปได้ทั้งหมดสำหรับแบบสอบของนักเรียนในระดับชั้นที่ต่ำกว่าและแบบสอบของนักเรียนในระดับชั้นที่สูงกว่าตามการศึกษาของ Yon (2007) ซึ่งเป็นค่าที่คาดว่าคลอบคลุมค่าที่เป็นไปได้ทั้งหมด ซึ่งรายละเอียดได้กล่าวถึงในบทที่ 3 ปรากฏในตารางที่ 3.3 และ 3.4

ขอบเขตของการวิจัย

1. การวิจัยครั้งนี้ใช้วิธีการศึกษาโดยการจำลองข้อมูล เพื่อศึกษาค่าพารามิเตอร์การแปลงคะแนนของข้อสอบร่วม จากระดับชั้นที่ 1 ไปยังระดับชั้นที่ 2 และจากระดับชั้นที่ 2 ไปยังระดับชั้นที่ 3 ซึ่งสถานการณ์การจำลองข้อมูลในครั้งนี้ เป็นไปตามเงื่อนไขหรือตัวแปรอิสระ 3 ตัว ได้แก่ 1) วิธีการเชื่อมโยงคะแนน 2 วิธี คือ วิธี M และ วิธี NOP 2) โครงสร้างมิติความสามารถ 2 ลักษณะ คือ โครงสร้างที่ไม่ซับซ้อน (Approximate Simple Structure; APSS) และโครงสร้างที่ซับซ้อนมากขึ้น(Mixed Structure; MS) และระดับความสัมพันธ์ระหว่างมิติความสามารถ 3 ระดับ คือ ไม่มีความสัมพันธ์ (r=0.0) มีความสัมพันธ์ระดับกลาง (r=0.5) และมีความสัมพันธ์ระดับสูง (r=0.8) ซึ่งแต่ละสถานการณ์จะทำซ้ำทั้งหมด 20 รอบ โดยมีคุณภาพของการเชื่อมโยงคะแนนเป็นตัวแปรตาม

2. โมเดลการวัดที่ใช้ในการวิจัยครั้งนี้เป็นโมเดลโลจิสติกแบบพหุมิติ ชนิด 3 พารามิเตอร์ (Multidimensional 3-Parameter Logistic Model; M3PL) ซึ่งเป็นโมเดล MIRT ที่การประมาณค่าความสามารถ (θ) ในมิติต่างกันสามารถชดเชยกันได้ (Compensatory Model) โดยมีการตอบสนองข้อสอบที่มีการตรวจให้คะแนน 2 ค่า และใช้เทคนิควิธีการสร้างสเกลแนวตั้งแบบพหุมิติ (Multidimensional Vertical Scaling) ตามโมเดล MIRT จากผู้สอบที่มีความสามารถแตกต่างกัน 3 ระดับชั้น ออกแบบการเก็บข้อมูลใช้รูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้แบบสอบร่วม (Anchor-Test Nonequivalent Group Design) ซึ่งมีข้อสอบร่วมชนิดภายใน (Internal-Common Items) จำนวน 30 ข้อ หรือ 30% ของจำนวนข้อสอบทั้งหมด 100 ข้อ

3. ดัชนีที่นำมาใช้ในการประเมินคุณภาพของการเชื่อมโยงคะแนนประกอบด้วยดัชนีรากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) และความลำเอียง (BIAS) โดยเปรียบเทียบค่าที่ได้ ถ้าเงื่อนไขที่ศึกษาใดมีค่าที่น้อยกว่าแสดงว่ามีคุณภาพในการเชื่อมโยงคะแนนได้ดีกว่า พร้อมทั้งค่าความสัมพันธ์ระหว่างคะแนนที่แปลงได้กับคะแนนเกณฑ์ที่ประมาณได้จากแบบสอบฐาน (CORR) โดยถ้ามีค่าสัมประสิทธิ์สหสัมพันธ์มากกว่าแสดงว่ามีคุณภาพของการเชื่อมโยงคะแนนได้ดีกว่า นอกจากนี้มีการวิเคราะห์ความแปรปรวนแบบพหุจำแนก 3 ทาง (Three-way MANOVA) เพื่อเปรียบเทียบค่าเฉลี่ยของ RMSE, BIAS และ CORR สำหรับศึกษาปฏิสัมพันธ์ของวิธีการเชื่อมโยงคะแนน โครงสร้างมิติความสามารถและระดับความสัมพันธ์ของมิติความสามารถ

ข้อจำกัดของการวิจัย

1. การวิจัยครั้งนี้ไม่ได้เจาะจงระดับชั้นในการศึกษาอย่างชัดเจน เป็นเพียงการจำลองข้อมูลที่แสดงให้เห็นถึงระดับขั้นของการเชื่อมโยงคะแนนที่ผู้สอบมีความสามารถต่างกันใน 3 ระดับ คือ ระดับที่ 1 แสดงถึงความสามารถขั้นต้น (Beginner) ระดับที่ 2 แสดงถึงความสามารถขั้นกลาง (Intermediate) และระดับที่ 3 แสดงถึงความสามารถขั้นสูง (Advance) โดยศึกษาการเชื่อมโยงคะแนนจากระดับที่ 1 ไปยังระดับที่ 2 และจากระดับที่ 2 ไปยังระดับที่ 3

2. จากผลการศึกษาพบว่าค่าความยากของข้อสอบ (d) ที่ได้จากการแปลงคะแนน ไม่แตกต่างกันระหว่างวิธี M และวิธี NOP ของแต่ละเงื่อนไขที่ศึกษา ดังนั้นในการนำเสนอผลการเปรียบเทียบคุณภาพของการเชื่อมโยงคะแนนในครั้งนี้จะเปรียบเทียบเฉพาะในบริบทของค่าพารามิเตอร์อำนาจจำแนกของข้อสอบในมิติที่ 1 (a_1) และ มิติที่ 2 (a_2) เท่านั้น

คำจำกัดความที่ใช้ในการวิจัย

การเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (MIRT Linking) หมายถึง กระบวนการในการปรับคุณลักษณะที่แตกต่างกันของข้อสอบ เช่น ความยากของข้อสอบ และอำนาจจำแนกของข้อสอบ เพื่อแปลงคะแนนของผู้สอบจากแบบสอบฉบับหนึ่งไปยังแบบสอบอีกฉบับหนึ่งให้สามารถเปรียบเทียบกันได้โดยตรง ซึ่งแบบสอบทั้งสองฉบับที่ทำการปรับเทียบกันมีลักษณะการวัด 2 มิติ ตามโมเดลการวิเคราะห์ของทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Item Response Theory; MIRT)

มิติของแบบสอบ (Dimensional of Test) หมายถึง ลักษณะของความเป็นตัวแทนทางทฤษฎีภายใต้คุณลักษณะแฝง (Trait) แนวความคิด (Concept) คุณลักษณะ (Attribute) กระบวนการ (Process) โครงสร้าง (Structure) โดยใช้แบบสอบซึ่งเป็นชุดของข้อสอบเป็นเครื่องมือวัดดังกล่าว ซึ่งโดยทั่วไปสามารถแบ่งลักษณะมิติของแบบสอบได้เป็น 2 ประเภทที่แตกต่างกัน คือ 1) แบบสอบที่มีลักษณะการวัดคุณลักษณะเด่นเพียงคุณลักษณะเดียว เรียกว่า “ความเป็นเอกมิติของแบบสอบ” และ 2) แบบสอบที่มีลักษณะการวัดคุณลักษณะเด่นได้หลายคุณลักษณะในคราวเดียวกัน เรียกว่า “ความเป็นพหุมิติของแบบสอบ” โดยในความหมายของการศึกษาในครั้งนี้ได้ใช้วิธีในการตรวจสอบมิติตามมโนทัศน์ของคุณลักษณะแฝง (Latent Trait Theory) โดยใช้การวิเคราะห์โครงสร้างคุณลักษณะแฝง (Latent Structure Analysis) ประกอบกับการวิเคราะห์เนื้อหา (Content Analysis) เพื่อเป็นแนวทางในการตรวจสอบมิติ ซึ่งมี 2 วิธี วิธีแรกอาจให้ผู้เชี่ยวชาญที่มีประสบการณ์ในการตรวจสอบหรือตัดสินมิติที่ต้องการวัด หรือวิธีที่สองการตัดสินโดยการยึดเนื้อหาเป็นหลัก

ความเป็นพหุมิติของแบบสอบ (Multidimensional of Test) หมายถึง ลักษณะของโครงสร้าง หรือคุณลักษณะภายในที่มีลักษณะเป็นองค์ประกอบเชิงซ้อน ภายใต้การวัดหลายมิติ ตัวอย่างเช่น แบบสอบคณิตศาสตร์ชุดหนึ่งอาจประกอบไปด้วยการวัดพีชคณิตและเรขาคณิต ในกรณีนี้ ชุดของแบบสอบย่อยมิติแรกกำหนดให้เป็นการวัดทางด้านพีชคณิต ส่วนมิติที่ 2 เป็นการวัดทางด้านเรขาคณิต ซึ่งเป็นการศึกษาในลักษณะ 2 มิติ เป็นต้น ซึ่งในการวิจัยนี้จะยึดนิยามตามมโนทัศน์ของคุณลักษณะแฝง (Latent Trait Theory) โดยใช้การวิเคราะห์โครงสร้างคุณลักษณะแฝง (Latent Structure Analysis) ซึ่งในโมเดลของคุณลักษณะแฝง สมมติให้ผู้เข้าสอบได้ตอบสนองข้อสอบที่มีการตรวจให้คะแนนแบบสองค่า (0-1) และพิสูจน์แล้วว่าข้อสอบเหล่านี้เป็นอิสระจากกันทางสถิติ ถ้ามีมากกว่า 1 คุณลักษณะแฝงที่สามารถอธิบายการแจกแจงของรูปแบบการตอบสนองของข้อสอบ n ข้อ แสดงว่าชุดของข้อสอบนี้มีความเป็นพหุมิติ ประกอบการพิจารณาความเป็นตัวแทนร่วมกับการกำหนดแบบแผนแบบสอบ (Test Specification) เพื่อใช้จำแนกมิติที่จะวัด

การสร้างสเกลแนวตั้งแบบพหุมิติ (Multidimensional Vertical Scaling) หมายถึง การเชื่อมโยงคะแนนระหว่างแบบสอบต่างฉบับกัน เมื่อแบบสอบต่างฉบับมุ่งวัดลักษณะเดียวกัน แต่ศึกษาใน 2 มิติ ซึ่งมีระดับความยากแตกต่างกันและกลุ่มผู้สอบมีการแจกแจงความสามารถอยู่คนละประชากร เพื่อปรับเทียบว่าคะแนนที่สอบได้จากฉบับหนึ่งเทียบเป็นเท่าไรของฉบับอื่นที่วัดต่างระดับกัน โดยแยกศึกษาในแต่ละมิติ ซึ่งการศึกษาในครั้งนี้มุ่งศึกษาใน 3 ระดับชั้น คือ ระดับที่ 1-3 โดยแบบสอบในระดับที่ 3 เป็นแบบสอบฐาน (Base Test) ส่วนแบบสอบในระดับชั้นที่ต่ำกว่าเพื่อนำมาเปรียบเทียบกับผลที่ได้กำหนดให้เป็นแบบสอบปรับเทียบคะแนน (Equated Test) จะถูกจำลองขึ้นตามการออกแบบการใช้ข้อสอบร่วมจากกลุ่มที่ไม่เท่าเทียมกัน (Common-Item Non-Equivalent Group Design) โดยการออกแบบในครั้งนี้ นักเรียนในแต่ละระดับชั้นจะทำ

แบบสอบในระดับชั้นของตนเองและแบบสอบในระดับชั้นสูงกว่าที่อยู่ติดกัน ซึ่งการจำลองข้อมูลจะมีการสร้างขึ้นโดยกำหนดให้นักเรียนในระดับที่ 1 (n=2,000 คน) ทำแบบสอบระดับที่ 2 ส่วนนักเรียนในระดับที่ 2 (n=2,000 คน) ทำแบบสอบในระดับตนเองและทำแบบสอบระดับที่ 3 ส่วนนักเรียนระดับที่ 3 ทำแบบสอบในระดับชั้นของตนเอง

แบบสอบปรับเทียบคะแนน (Equated Test) หมายถึง แบบวัดในระดับชั้นที่ต่ำกว่าที่จะนำไปสู่การแปลงคะแนนในระดับชั้นที่สูงกว่า ซึ่งในการวิจัยครั้งนี้เป็นแบบสอบในระดับที่ 1 และ 2

แบบสอบฐาน (Base Test) หมายถึง แบบสอบในระดับชั้นที่อยู่สูงกว่า โดยในการวิจัยครั้งนี้เป็นแบบสอบระดับที่ 3 ซึ่งเป็นแบบสอบในระดับชั้นสูงสุดที่ทำการศึกษาในการจำลองข้อมูล

ข้อสอบร่วม (Common Items) หมายถึง ข้อสอบร่วมระหว่างชั้นที่ต่อเนื่องที่อยู่ในแบบสอบร่วมภายในแต่ละระดับชั้นที่ต่อเนื่องนั้น แบ่งออกเป็นสองชุด ดังนี้

ข้อสอบร่วมระหว่างชั้นที่ต่อเนื่อง ระดับที่ 1-2 จำนวน 30 ข้อ จัดอยู่ในแบบสอบปรับเทียบคะแนนของระดับที่ 1 ไประดับที่ 2

ข้อสอบร่วมระหว่างชั้นที่ต่อเนื่อง ระดับที่ 2-3 จำนวน 30 ข้อ จัดอยู่ในแบบสอบปรับเทียบคะแนนของระดับที่ 2 ไประดับที่ 3

คะแนนสมมูล (Equivalent Score) หมายถึง คะแนนระหว่างแบบสอบต่างชุดที่เทียบคะแนนกันได้ว่าเป็นคะแนนที่เทียบเท่ากัน

วิธีการเชื่อมโยงคะแนน (Linking Method) หมายถึง กระบวนการปรับคุณลักษณะที่แตกต่างกันของข้อสอบ เช่น ความยากของข้อสอบ และอำนาจจำแนกของข้อสอบ เพื่อแปลงคะแนนจากแบบสอบปรับเทียบคะแนนไปยังแบบสอบฐาน ซึ่งในการวิจัยครั้งนี้มี 2 วิธี จำแนกตามลักษณะการหมุนแกน คือ 1) วิธีการหมุนแกนแบบตั้งฉาก (Orthogonal Procrustes Method) โดยวิธี M (Min, 2003) และ 2) วิธีการหมุนแกนแบบไม่ตั้งฉาก (Non-Orthogonal Procrustes Method) โดยวิธี NOP (Reckase & Martineau, 2004)

การหมุนแกนแบบตั้งฉาก (Orthogonal Procrustes Method) หมายถึง กระบวนการที่ใช้ในการแปลงคะแนนจากแบบสอบปรับเทียบคะแนนไปยังแบบสอบฐาน เพื่อประมาณค่าพารามิเตอร์ของข้อสอบตามโมเดล MIRT ที่มีการกำหนดระบบแกนอ้างอิง (Reference System) หรือมิติของ θ ให้อยู่ในตำแหน่งที่เหมาะสม (Scale Indeterminacy) โดยใช้วิธี M (Min's Method) (Min, 2003) ซึ่งแทนค่าพารามิเตอร์การเลื่อนแกนให้อยู่ในรูปของเมตริกซ์ (Diagonal Dilation Matrix) ยึดการหมุนแกนและระยะของมาตรวัดที่ระยะทำมุม 90°

การหมุนแกนแบบไม่ตั้งฉาก (Non-Orthogonal Procrustes Method) หมายถึง กระบวนการที่ใช้ในการแปลงคะแนนจากแบบสอบปรับเทียบคะแนนไปยังแบบสอบฐาน เพื่อประมาณค่าพารามิเตอร์ของข้อสอบตามโมเดล MIRT ที่มีการกำหนดระบบแกนอ้างอิง (Reference System) หรือมิติของ θ ให้อยู่ในตำแหน่งที่เหมาะสม (Scale Indeterminacy) โดยใช้วิธี NOP (Non-Orthogonal Procrustes Method) (Reckase & Martineau, 2004) ซึ่งไม่พิจารณา

ค่าพารามิเตอร์การเลื่อนแกนทั้งในรูปสเกลลาและเมตริกซ์ ยึดการหมุนแกนและระยะของมาตรวัดเป็นมุมแหลม (Oblique Procrustes)

โครงสร้างมิติความสามารถ (Dimensional Structure) หมายถึง ลักษณะที่สะท้อนถึงน้ำหนักเนื้อหาของข้อสอบแต่ละข้อกับความสัมพันธ์ของเนื้อหาในมิติที่ 1 และ 2 ซึ่งสามารถพิจารณาได้จากการกำหนดแบบแผนข้อสอบ (Test Blueprint) ของน้ำหนักข้อสอบแต่ละข้อกับเนื้อหาที่ศึกษาในแต่ละมิติ สำหรับการวิจัยครั้งนี้คือลักษณะของความสัมพันธ์ระหว่างเวคเตอร์ข้อสอบแต่ละข้อกับมิติที่ศึกษา จะพิจารณาตามขนาดความซับซ้อนของโครงสร้าง 2 ลักษณะ โดยโครงสร้างแรกเป็นโครงสร้างที่ไม่ซับซ้อน (Approximate Simple Structure; APSS) ส่วนโครงสร้างที่ 2 เป็นโครงสร้างที่มีความซับซ้อนมากขึ้น (Mixed Structure; MS)

โครงสร้างที่ไม่ซับซ้อน (Approximate Simple Structure; APSS) หมายถึง ลักษณะของชุดข้อสอบที่แต่ละข้อมีความสัมพันธ์กันสูงในมิติที่หนึ่งมากกว่ามิติที่สอง นั่นคือกลุ่มของข้อสอบจำนวนหนึ่งมีค่าอำนาจจำแนกที่สูงในมิติเดียวกันแต่มีค่าอำนาจจำแนกต่ำสำหรับอีกมิติหนึ่ง ในการศึกษาครั้งนี้ผู้วิจัยสนใจศึกษาแบบสอบที่มีลักษณะ 2 มิติ โดยมีข้อสอบร่วมในแต่ละระดับทั้งหมด 30 ข้อ โดยในแต่ละมิติมีจำนวนข้อสอบเท่ากันมิติละ 15 ข้อ ลักษณะเวคเตอร์ของข้อสอบในมิติที่ 1 (θ_1) จะทำมุมกับแกนของมิติที่ 1 ตั้งแต่ 0°–15° ($0^\circ \leq \alpha_k \leq 15^\circ$) เมื่อ α_k เป็นมุมระหว่างเวคเตอร์ในมิติที่ 1 กับแกนของมิติที่ 1 ในขณะที่ลักษณะเวคเตอร์ของข้อสอบในมิติที่ 2 (θ_2) จะทำมุมกับแกนของมิติที่ 1 ช่วงมุมตั้งแต่ 75°–90° (90-α_k)

โครงสร้างที่ซับซ้อน (Mixed Structure; MS) หมายถึง ลักษณะของชุดข้อสอบที่แต่ละข้อมีความสัมพันธ์กันทั้งในมิติที่หนึ่งและมิติที่สอง โดยจะมีการวัดในแต่ละมิติและวัดมิติโดยรวม (Composites of Dimensions) (Kim, 1994) ซึ่งในการศึกษาครั้งนี้จะศึกษาเช่นเดียวกับโครงสร้างที่ไม่ซับซ้อน แต่ขนาดของมุมระหว่างข้อสอบกับแกนของมิติจะแตกต่างกันไป โดยมีแบบสอบทั้งหมด 4 ชุดย่อย แบบสอบย่อย 2 ชุดแรก ชุดละ 7 ข้อ จะแยกเป็นมิติที่ 1 (0°–15°) ประกอบไปด้วยข้อที่ 1-7 และ มิติที่ 2 (75°–90°) ซึ่งประกอบไปด้วยข้อที่ 23-30 ส่วนแบบสอบย่อยอีก 2 ส่วนที่เหลือ คือ ข้อที่ 8-15 และ 16-22 จะวัดในลักษณะของมิติรวม ข้อที่ 8-15 ข้อสอบมีลักษณะใกล้เคียงกับมิติที่ 1 มากกว่ามิติที่ 2 ในขณะที่ข้อ 16-22 ข้อสอบมีลักษณะใกล้เคียงกับมิติที่ 2 มากกว่ามิติที่ 1 โดยการศึกษาครั้งนี้ได้นำแนวคิดของ Wang (1986) ที่มีการนำเสนอแนวคิดแกนโดยรวม (Reference Composite) ในลักษณะของการศึกษาแบบพหุมิติ (Multivariate Space) ซึ่งค่าที่ได้สามารถสะท้อนถึงค่าอำนาจจำแนกรวม (MDISCs) และค่าความแปรปรวนความแปรปรวนร่วมของการกระจายความสามารถที่มีลักษณะแบบหลายมิติ (Ackerman, 1992)

ความสัมพันธ์ระหว่างมิติความสามารถ หมายถึง ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าการแจกแจงความสามารถของผู้สอบในมิติที่ 1 กับค่าการแจกแจงความสามารถของผู้สอบในมิติที่ 2 เช่น ผู้สอบที่มีความสามารถทางด้านการคำนวณ จำเป็นต้องอาศัยความสามารถทางด้านภาษาซึ่งสะท้อนความสามารถทางด้านคำนวณและภาษามีความสัมพันธ์กัน เป็นต้น ถ้าค่าสัมประสิทธิ์

สหสัมพันธ์มีค่าเท่ากับ 0 แสดงว่าการประมาณค่าความสามารถของผู้สอบในมิติที่ 1 กับมิติที่ 2 เป็นอิสระต่อกัน แต่ถ้าค่าสัมประสิทธิ์สหสัมพันธ์มีค่าไม่เท่ากับ 0 แสดงว่าการประมาณค่าความสามารถของผู้สอบในมิติที่ 1 กับมิติที่ 2 มีความเกี่ยวข้องสัมพันธ์กัน โดยในการศึกษาครั้งนี้ผู้วิจัยสนใจศึกษาความสัมพันธ์ 3 ระดับ ประกอบด้วย ไม่มีความสัมพันธ์ (r=0.0) มีความสัมพันธ์ระดับกลาง (r=0.5) และมีความสัมพันธ์ระดับสูง (r=0.8)

คุณภาพของการเชื่อมโยงคะแนน (Quality of Linking Method) หมายถึง ความถูกต้องแม่นยำของการเชื่อมโยงคะแนน ซึ่งพิจารณาจากดัชนีที่พิจารณาขนาดความคลาดเคลื่อนที่เกิดจากวิธีการเชื่อมโยงคะแนน โดยคำนวณค่าความแตกต่างระหว่างการประมาณจากแบบสอบฐาน (Base Estimates) และการประมาณค่าการแปลงคะแนน (Transformed Estimates) ที่ได้จากแบบสอบปรับเทียบคะแนน (Equated Estimates) รวมถึงการพิจารณาสหสัมพันธ์ระหว่างการประมาณค่าพารามิเตอร์จากแบบสอบฐานและค่าการแปลงคะแนนที่ได้จากแบบสอบปรับเทียบคะแนน ซึ่งดัชนีที่นำมาใช้ในการเปรียบเทียบคุณภาพของการเชื่อมโยงคะแนนในครั้งนี้ ประกอบด้วยดัชนีค่าความลำเอียง (BIAS) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) โดยเปรียบเทียบค่าที่ได้ ถ้าเงื่อนไขที่ศึกษาใดมีค่าที่น้อยกว่าแสดงว่ามีคุณภาพในการเชื่อมโยงคะแนนได้ดีกว่า พร้อมทั้งเปรียบเทียบค่าค่าสัมประสิทธิ์สหสัมพันธ์ภายหลังจากการแปลงคะแนนของแบบสอบปรับเทียบคะแนน (Equated Test) กับแบบสอบฐาน (Base Test) โดยถ้ามีค่าความสัมพันธ์มากกว่าแสดงว่ามีคุณภาพของการเชื่อมโยงคะแนนได้ดีกว่า พร้อมทั้งวิเคราะห์ MANOVA เพื่อเปรียบเทียบค่าเฉลี่ยของดัชนี RMSE, BIAS และ ค่าสัมประสิทธิ์สหสัมพันธ์ สำหรับประเมินอิทธิพลของวิธีการเชื่อมโยงคะแนนและโครงสร้างมิติความสามารถที่มีต่อคุณภาพของการเชื่อมโยงคะแนน

โดยมีรายละเอียดของดัชนีที่ใช้ในการประเมินคุณภาพของการเชื่อมโยงคะแนน ดังนี้

(1) ค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (Root-Mean Square Error; RMSE) หมายถึง ขนาดของความแตกต่างระหว่างการประมาณค่าจากแบบสอบปรับเทียบคะแนนและแบบสอบฐาน นั่นคือ มีความแตกต่างทั้งหมดของการแปลงคะแนนข้ามกลุ่มข้อสอบร่วม (Common Items) มากน้อยเพียงใด ซึ่งดัชนี RMSE สะท้อนให้เห็นถึงความคงที่ (Stability) ของการเชื่อมโยงคะแนน โดยมีหลักเกณฑ์ในการพิจารณาจากการเปรียบเทียบค่าที่ได้ ถ้ามีค่าดัชนี RMSE น้อยกว่า แสดงว่ามีความคงที่ของการเชื่อมโยงคะแนนสูงกว่า สะท้อนให้เห็นถึงความคลาดเคลื่อนของการเชื่อมโยงคะแนนมีน้อย แต่ถ้ามีค่าดัชนี RMSE มากกว่า แสดงว่ามีความคงที่ของการเชื่อมโยงคะแนนต่ำสะท้อนให้เห็นถึงความคลาดเคลื่อนของการเชื่อมโยงคะแนนมีมาก ซึ่งงานวิจัยครั้งนี้ทำการเปรียบเทียบความแตกต่างระหว่างค่าเฉลี่ยจาก Ln RMSE ซึ่งเป็นค่าที่ผกผันกับ RMSE นั่นคือถ้ามีค่า Ln RMSE มาก จะมีค่า RMSE น้อย แสดงถึงความคงที่ของการเชื่อมโยงคะแนน แต่ถ้ามีค่า Ln RMSE น้อย จะมี

ค่า RMSE มาก แสดงถึงความคลาดเคลื่อนของการเชื่อมโยงคะแนน ซึ่งสามารถแสดงได้ดังสมการที่ 1.1

(2) ค่าความลำเอียง (BIAS) หมายถึง ค่าความคลาดเคลื่อนในการเชื่อมโยงคะแนน นั่นคือ การประมาณค่าที่ได้จากการแปลงคะแนนของแบบสอบปรับเทียบคะแนนมีผลต่างจากค่าพารามิเตอร์ที่แท้จริงของข้อสอบโดยเฉลี่ยมากน้อยเพียงใด ซึ่งดัชนี BIAS สะท้อนให้เห็นถึงความถูกต้อง (Accuracy) ของการเชื่อมโยงคะแนน โดยมีหลักเกณฑ์ในการพิจารณาจากการเปรียบเทียบค่าที่ได้ ถ้ามีค่าดัชนี BIAS ใกล้ศูนย์มากกว่าสะท้อนให้เห็นถึงการประมาณค่าใกล้เคียงกับค่าที่แท้จริงมากกว่า โดยเครื่องหมายที่ได้จะเป็นตัวสะท้อนถึงความคลาดเคลื่อนของการประมาณค่าที่ห่างจากค่าที่แท้จริงในลักษณะใด นั่นคือ ถ้าดัชนี BIAS ติดลบ (-) แสดงว่ามีการประมาณค่าต่ำกว่าความเป็นจริง (Underestimate) ถ้าเป็นบวก (+)แสดงว่ามีการประมาณค่าสูงกว่าความเป็นจริง (Overestimate) ซึ่งสามารถแสดงได้ดังสมการที่ 1.2

$$RMSE = \left[\frac{\sum_{i=1}^{I} \left(a_{1i}^{*} - a_{1i} \right)^2}{I - 1} \right]^{1/2} \quad \text{.....(1.1)}$$

$$BIAS = \sum_{i=1}^{I} \frac{\left(a_{1i}^{*} - a_{1i} \right)}{I} \quad \text{.....(1.2)}$$

เมื่อ a_{1i} แทน ค่าพารามิเตอร์อำนาจจำแนก (Discrimination Parameter) ของมิติที่ 1 สำหรับข้อสอบข้อที่ i

a^{*}_{1i} แทน ค่าอำนาจจำแนกที่ได้จากการแปลงคะแนน (Transformed Discrimination)

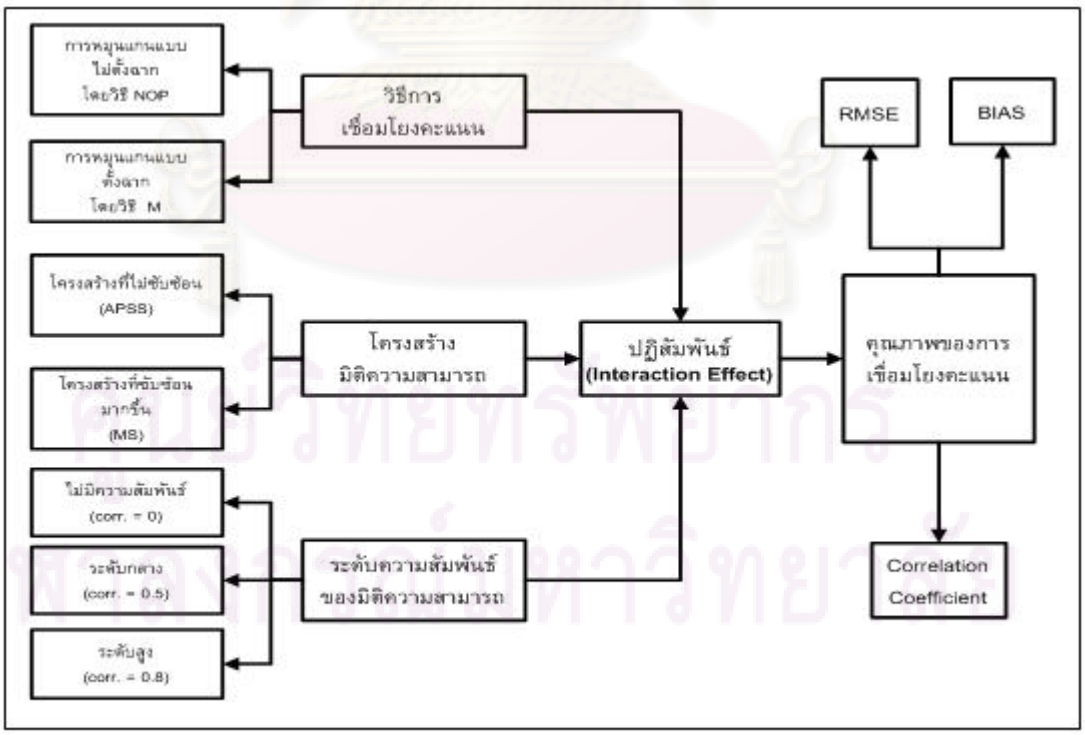
I แทน จำนวนของข้อสอบร่วม

(3) ค่าความสัมพันธ์ระหว่างคะแนนที่แปลงได้กับคะแนนเกณฑ์ (CORR) หมายถึง ค่าที่แสดงความสัมพันธ์ระหว่างคะแนนที่แปลงได้จากแบบสอบปรับเทียบคะแนนกับคะแนนที่ประมาณได้จากแบบสอบฐาน ซึ่งค่า CORR สะท้อนให้เห็นถึงความสัมพันธ์ระหว่างค่าที่แปลงได้กับคะแนนเกณฑ์ โดยมีหลักเกณฑ์ในการพิจารณาจากการเปรียบเทียบค่าที่ได้ ถ้ามีค่า CORR สูง จะสะท้อนให้เห็นถึงคุณภาพของการเชื่อมโยงคะแนนที่ดี

ปฏิสัมพันธ์ของวิธีการเชื่อมโยงคะแนนและเงื่อนไขที่ทำการศึกษาต่อคุณภาพของการเชื่อมโยงคะแนน หมายถึง ค่าที่ได้จากการวิเคราะห์ความแปรปรวนแบบพหุ (MANOVA) เพื่อใช้ในการศึกษาปฏิสัมพันธ์ของวิธีการเชื่อมโยงคะแนนทั้งสองวิธี และเงื่อนไขที่ทำการศึกษา คือ โครงสร้างมิติความสามารถที่แตกต่างกัน 2 ลักษณะ และระดับความสัมพันธ์ของมิติความสามารถที่แตกต่างกัน 3 ระดับ ที่ส่งผลต่อคุณภาพของการเชื่อมโยงคะแนน 3 ชนิด คือ ดัชนี RMSE BIAS และ CORR

กรอบแนวคิดที่ใช้ในการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนน 2 วิธี ระหว่างการหมุนแกนแบบตั้งฉากโดย วิธี M (Min's Method) และการหมุนแกนแบบไม่ตั้งฉากโดยวิธี NOP (Non-Orthogonal Procrustes) โดยเงื่อนไขในการศึกษา คือ ระดับความสัมพันธ์ระหว่างมิติความสามารถ 3 ระดับ ประกอบด้วย ไม่มีความสัมพันธ์ (r=0.0) มีความสัมพันธ์ระดับกลาง(r=0.5) และมีความสัมพันธ์ระดับสูง (r=0.8) ที่มีโครงสร้างมิติความสามารถที่แตกต่างกัน 2 ลักษณะระหว่างโครงสร้างที่ไม่ซับซ้อน (Approximate Simple Structure; APSS) และโครงสร้างที่ซับซ้อนมากขึ้น (Mixed Structure; MS) เกณฑ์ที่ใช้ในการเปรียบเทียบคุณภาพวิธีการเชื่อมโยงคะแนนทั้ง 2 วิธี ทำการพิจารณาจากดัชนีค่าความลำเอียง (BIAS) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (Root-Mean Square Error; RMSE) พร้อมทั้งศึกษาสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) ของผลที่ได้จากการแปลงคะแนนระหว่างแบบสอบฐาน และแบบสอบปรับเทียบคะแนน พร้อมทั้งวิเคราะห์ความแปรปรวนแบบพหุ (MANOVA) เพื่อเปรียบเทียบค่าเฉลี่ยของ BIAS, RMSE และ CORR สำหรับศึกษาปฏิสัมพันธ์ของวิธีการเชื่อมโยงคะแนน โครงสร้างมิติความสามารถและระดับความสัมพันธ์ของมิติความสามารถ เพื่อประโยชน์ในการนำไปเลือกใช้กับสถานการณ์ต่าง ๆ ให้เหมาะสมกับข้อมูลจริงต่อไปในอนาคต โดยสามารถแสดงกรอบแนวคิดในการวิจัยดังแผนภาพที่ 1.1



แผนภาพที่ 1.1 กรอบแนวคิดในการวิจัย

ประโยชน์ที่คาดว่าจะได้รับ

1. ประโยชน์ทางด้านวิชาการ

1.1 ได้ขยายองค์ความรู้ในด้านการเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (MIRT Linking) ทั้งทางทฤษฎีและการปฏิบัติเพื่อนำไปประยุกต์ใช้ในการศึกษาพัฒนาการภาคตัดขวางโดยการสร้างสเกลแนวตั้งที่มี 2 มิติ โดยเฉพาะแบบสอบวัดผลสัมฤทธิ์ทางการเรียน

1.2 ผลการวิจัยให้ข้อมูลสารสนเทศเกี่ยวกับผลการเปรียบเทียบคุณภาพของการเชื่อมโยงคะแนนที่มีวิธีการแตกต่างกัน 2 วิธี ระหว่างการหมุนแกนแบบตั้งฉากกับไม่ตั้งฉาก ในเงื่อนไขของการศึกษา คือ โครงสร้างมิติความสามารถ และระดับความสามารถของมิติที่แตกต่างกัน ทำให้ได้ผลสรุปที่ชัดเจนขึ้นเกี่ยวกับวิธีการเชื่อมโยงคะแนนในสถานการณ์การศึกษาที่แตกต่างกัน และเป็นพื้นฐานที่สำคัญในการวิจัยเกี่ยวกับการวัดและประเมินผล โดยเฉพาะอย่างยิ่งการศึกษาการวัดพัฒนาภาคตัดขวางที่มีหลายมิติ

2. ประโยชน์ทางด้านการนำไปใช้

2.1 ผลการวิจัยสามารถนำไปใช้ประกอบการตัดสินใจในการเลือกใช้วิธีการเชื่อมโยงคะแนนที่เหมาะสมและสอดคล้องกับสถานการณ์ที่ต้องการศึกษามากที่สุด เพื่อนำมาศึกษาพัฒนาการของผู้เรียนโดยศึกษาในแต่ละมิติและศึกษามิติโดยรวม

2.2 ผลการวิจัยเป็นแนวทางในการนำไปใช้วัดผลสัมฤทธิ์ทางการเรียนที่มีลักษณะการบูรณาการความรู้ข้ามเนื้อหาได้ทุกระดับช่วงชั้น และทุกกลุ่มสาระการเรียนรู้เพื่อวัดพัฒนาการการเรียนรู้ของผู้เรียนจากระดับชั้นหนึ่งไปอีกระดับชั้นหนึ่งที่มีลักษณะภาคตัดขวาง (Cross-sectional) ซึ่งเป็นการศึกษาในภาคการศึกษาหรือปีการศึกษาเดียวกัน โดยเฉพาะกลุ่มสาระการเรียนรู้คณิตศาสตร์และวิทยาศาสตร์ที่มีกลุ่มของเนื้อหาค่อนข้างเด่นชัด เช่น ในกลุ่มสาระคณิตศาสตร์ประกอบด้วย 5 กลุ่มที่สำคัญ คือ 1) จำนวนและการดำเนินการ 2) การวัด 3) เรขาคณิต 4) พีชคณิต และ 5) การวิเคราะห์ข้อมูลและความน่าจะเป็น สำหรับกรณีที่ศึกษาในรายวิชาวิทยาศาสตร์สามารถแยกเป็นกลุ่มเนื้อหาที่สำคัญ คือ วิทยาศาสตร์ชีววิทยา (Biology Science) วิทยาศาสตร์กายภาพ (Physical Science) และวิทยาศาสตร์สิ่งแวดล้อม (Environmental Science) เป็นต้น ซึ่งจากที่กล่าวมาสามารถศึกษาได้คราวละ 2 มิติ โดยทำการตรวจสอบความเป็นพหุมิติจากหลักฐานทางสถิติโดยการวิเคราะห์ตามโมเดล MIRT ร่วมกับการกำหนดแบบแผนของแบบสอบ (Test Specification) เพื่อเป็นแนวทางในการเขียนข้อสอบตามเนื้อหาและหน่วยที่ต้องการวัดที่แสดงถึงความเป็นตัวแทนของมิติที่มุ่งวัดได้อย่างถูกต้องและการวิเคราะห์เนื้อหา (Content Analysis) เพื่อเป็นแนวทางในการตรวจสอบมิติ ซึ่งมี 2 วิธี วิธีแรกอาจให้ผู้เชี่ยวชาญที่มีประสบการณ์ในการตรวจสอบหรือตัดสินมิติที่ต้องการวัด หรือวิธีที่สองการตัดสินโดยการยึดเนื้อหาเป็นหลัก

บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

ผู้วิจัยได้ทบทวนวรรณคดีที่เกี่ยวข้องกับวิธีการเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Item Response Theory Linking) และการสร้างสเกลแนวตั้งแบบพหุมิติ (Multidimensional Vertical Scaling) จากวารสาร Journal of Education Measurement, Applied Psychological Measurement, Applied Measurement in Education, Educational Measurement: Issues and Practice, Journal of Applied Measurement และ Psychometrika โดยเป็นวารสารทางการวัดผลการศึกษาและจิตวิทยา ได้เผยแพร่ประเด็นสำคัญของหัวข้อการศึกษาวิธีการเชื่อมโยงคะแนนตามโมเดล MIRT การศึกษาหนังสืออีก 5 เล่มที่เกี่ยวข้องเป็นพิเศษ (Reckase, 1997; McDonald, 1999; Embretson & Reise, 2000; Kolen & Brennan, 2004; ศิริชัย กาญจนวาสี, 2545) รวมถึงการสืบค้นวารสารทางอิเล็กทรอนิคส์โดยการค้นคว้าทางอินเตอร์เน็ท ซึ่งประเด็นจากการศึกษาค้นคว้าในครั้งนี้ แบ่งการนำเสนอเอกสารและงานวิจัยที่เกี่ยวข้องออกเป็น 3 ตอน คือ

ตอนที่ 1 โมเดลการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Item Response Theory Model)

ตอนที่ 2 การเชื่อมโยงคะแนนตามโมเดลทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Item Response Theory Linking) และงานวิจัยที่เกี่ยวข้อง

ตอนที่ 3 การสร้างสเกลแนวตั้งพหุมิติ (Multidimensional Vertical Scaling) และงานวิจัยที่เกี่ยวข้อง

โดยมีรายละเอียดในแต่ละตอนดังต่อไปนี้

ตอนที่ 1 โมเดลการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Item Response Theory Model)

โมเดลการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Item Response Theory Model; MIRT Model) ถือได้ว่าเป็นแนวคิดที่แผ่ขยายมาจากโมเดลการตอบสนองข้อสอบแบบพหุมิติ (Unidimensional Item Response Theory Model; UIRT Model) เพื่อแก้ไขจุดอ่อนของข้อตกลงเบื้องต้นเกี่ยวกับความเป็นเอกมิติ โดยจะแตกต่างจากแนวคิดแบบเดิมในส่วนของระเบียบวิธีของโมเดล MIRT ที่นำมาประยุกต์และการแปลความหมายของผลที่ได้ ในที่นี้ผู้วิจัยได้นำเสนอรายละเอียดที่มีส่วนเกี่ยวข้องต่อการนำไปพัฒนาวิธีการเชื่อมโยงคะแนนแบบพหุมิติ ในประเด็นของมโนทัศน์เบื้องต้นเกี่ยวกับโมเดลการตอบสนองข้อสอบแบบพหุมิติ (MIRT Model) โดยมีรายละเอียดดังต่อไปนี้

1.1 มโนทัศน์เบื้องต้นเกี่ยวกับโมเดลการตอบสนองข้อสอบแบบพหุมิติ (MIRT Model)

1.1.1 ที่มาและความสำคัญของโมเดล MIRT

แนวคิดของทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (MIRT) ได้มาจากแนวคิดพื้นฐานที่สำคัญ จำแนกออกเป็น 2 กลุ่มแนวคิด โดยกลุ่มแรกเป็นการศึกษาโดย Spearman (1927) & Thurstone (1947) ทำการศึกษาและพบว่าโมเดล MIRT มีที่มาจากการวิเคราะห์องค์ประกอบ (Factor Analysis) เนื่องจากลักษณะของการวิเคราะห์องค์ประกอบจะมีความสัมพันธ์โดยตรงกับการวิเคราะห์ MIRT ส่วนอีกกลุ่มหนึ่ง เป็นการศึกษาของ Lazarsfeld (1950) กล่าวว่า MIRT ได้รับอิทธิพลมาจาก UIRT เช่นเดียวกับการวิเคราะห์องค์ประกอบ แต่แนวคิดเกี่ยวกับ UIRT ในช่วงแรกยังไม่เป็นที่นิยมเหมือนเช่นปัจจุบัน จากทั้งสองแนวคิดอาจกล่าวว่า MIRT เป็นแนวคิดที่แผ่ขยายมาจาก UIRT หรือเป็นกรณีเฉพาะของการวิเคราะห์องค์ประกอบ หรือโมเดลสมการเชิงโครงสร้าง (Structural Equation Modeling) นั่นเอง เพื่อให้เห็นถึงความเหมือนและความต่างของการวิเคราะห์องค์ประกอบซึ่งเป็นรากฐานของโมเดล MIRT ผู้วิจัยได้นำเสนอวิธีการวิเคราะห์องค์ประกอบและทฤษฎีการตอบสนองข้อสอบ (Item Response Theory; IRT) ที่มีความสัมพันธ์กับรากฐานของแนวคิด MIRT พร้อมทั้งเปรียบเทียบวิธีการของการวิเคราะห์องค์ประกอบ และ MIRT โดยมีรายละเอียดต่อไปนี้

(1) การวิเคราะห์องค์ประกอบ (Factor Analysis)

จากการศึกษาของ Reckase (1997) พบว่า มีนักวิชาหลายท่านได้มีการศึกษาและเห็นตรงกันว่าถึงแม้กระบวนการทางคณิตศาสตร์ของการวิเคราะห์องค์ประกอบและ MIRT จะมีลักษณะระเบียบวิธีทางด้านตัวเลขที่คล้ายกัน ซึ่งทั้งสองวิธีพยามที่จะระบุสเกลสมมติฐานเพื่อถอดแบบข้อมูลเพื่อนำไปสู่การวิเคราะห์ข้อมูล โดยสเกลของการวิเคราะห์องค์ประกอบและ MIRT จะต้องมีการกำหนดจุดกำเนิด (Origin) และหน่วยของการวัด (Unit of Measurement) เหมือนกัน แต่สิ่งที่ MIRT แตกต่างจากการวิเคราะห์องค์ประกอบ เมื่อพิจารณาถึงคุณลักษณะของข้อสอบหรือตัวแปรตั้งต้น (Input Variable) พบว่าการวิเคราะห์องค์ประกอบจะพิจารณาถึงความแตกต่างของลักษณะตัวแปรตั้งต้น เช่น ความแตกต่างของค่าเฉลี่ย ($\bar{x}$) และส่วนเบี่ยงเบนมาตรฐาน (SD) เป็นต้น รวมถึงความเที่ยงของการวิเคราะห์ข้อมูล เช่น การกำจัดตัวแปรแทรกซ้อนหรือตัวแปรที่ไม่จำเป็นออกจากโมเดลการวิเคราะห์ โดยการปรับค่าสถิติให้เป็นมาตรฐาน จะเห็นได้ว่าการวิเคราะห์องค์ประกอบจะให้ความสำคัญกับเมตริกซ์ความสัมพันธ์แต่ไม่ให้ความสำคัญกับความแตกต่างของค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของตัวแปรที่ทำการศึกษา เช่น จากการศึกษาของ Herman (1976) ได้อธิบายลักษณะของการวิเคราะห์องค์ประกอบว่า “หลักการเบื้องต้นของการวิเคราะห์องค์ประกอบคือ เป็นกระบวนการการจัดกระทำกับข้อมูลหรือตัวแปรเพื่อจัดกลุ่มตัวแปรในการจำแนกประเภทหรือองค์ประกอบให้ได้จำนวนน้อยที่สุด วิธีการดังกล่าวสามารถทำได้โดยการศึกษาความสัมพันธ์ระหว่างตัวแปร” ซึ่งเมตริกซ์ความสัมพันธ์ ถือเป็นข้อมูลหลักในการวิเคราะห์ โดยไม่ได้ให้ความสนใจกับค่าพารามิเตอร์ของข้อสอบที่เป็นคุณลักษณะที่มีความสัมพันธ์กับข้อสอบหรือ

ตัวแปรตั้งต้นโดยตรง เช่น ค่าความยากของข้อสอบ อำนาจจำแนกของข้อสอบ หรือโอกาสในการเดาของข้อสอบ เป็นต้น แต่จะพิจารณาตัวแปรแทรกซ้อนอื่นๆ ที่จะตัดออกจากโมเดลการวิเคราะห์ เพื่อให้ความสัมพันธ์มีความถูกต้องมากขึ้น (Carroll, 1945)

จากการศึกษาที่ผ่านมาพบว่า มีบุคคล 5 กลุ่ม ที่ศึกษาเกี่ยวกับการวิเคราะห์องค์ประกอบได้เด่นชัดที่สุด และเป็นการจุดประกายการแผ่ขยายแนวคิดของโมเดล MIRT ตามลำดับ (Reckase, 1997) ดังต่อไปนี้

1) Horst (1965) ถือว่าเป็นบุคคลแรกที่จุดประกายเกี่ยวกับแนวความคิดของ MIRT โดยทำการศึกษาการวิเคราะห์องค์ประกอบที่สามารถสรุปในรูปของเมตริกซ์ข้อมูล (Factor Analysis of Data Metrices) ซึ่งให้ความสำคัญกับความสอดคล้องของเมตริกซ์ข้อมูลเต็มรูปแบบ (Full Data Matrix) มากกว่าที่จะให้ความสำคัญกับเมตริกซ์ความสัมพันธ์ (Correlation Matrix) จากกลุ่มของตัวแปรสมมติฐาน โดยเขากล่าวว่า *"เราควรที่จะให้ความสำคัญกับตัวแปรทั้งหมดของการวิเคราะห์องค์ประกอบซึ่งไม่สนใจเฉพาะตัวแปรที่ต้องการศึกษาเท่านั้น แต่ต้องทำการพิจารณาทั้งเมตริกซ์ของคะแนนที่สังเกตได้ และเมตริกซ์ของคะแนนจริง ซึ่งโดยปกติแล้วตัวแปร (Treatment) ที่จะนำไปสู่การวิเคราะห์ ได้มาจากเมตริกซ์ความสัมพันธ์ของคะแนนที่สังเกตได้ วิธีการดังกล่าวเป็นแนวคิดที่จะนำไปสู่มโนทัศน์ที่คลาดเคลื่อน คิดว่าน่าจะให้สารสนเทศที่มากกว่าเมตริกซ์ข้อมูล แต่ความเป็นจริงแล้วควรที่จะให้ความสำคัญกับเมตริกซ์ข้อมูล เป็นอันดับแรกเพื่อที่จะได้พิจารณาถึงปัญหาและเทคนิคในการนำมาใช้วิเคราะห์องค์ประกอบ"*

โดยจากการพิจารณาเมตริกซ์ข้อมูลของคะแนนสังเกตได้ Horst ได้เปรียบเทียบความสัมพันธ์กับคุณลักษณะของตัวแปร โดยมุ่งประเด็นไปยังจุดเริ่มต้น (Origin) และหน่วยของการวัด (Unit of Measurement) และผลของการแปลงคะแนนในสเกลที่ได้จากการวิเคราะห์องค์ประกอบ โดยมุ่งไปที่เมตริกซ์แบบแบ่งสอง (Binary Matrices) และพยายามที่จะทำให้เป็นมาตรฐานของตัวแปรแบบแบ่งสอง นอกจากนี้ยังได้แนะนำเกี่ยวกับความแปรผันบางส่วนของความยากของข้อสอบ (Dispersion of Item Preference) ซึ่งเรียกกระบวนการนี้ว่า "Partialing Out the Simplex" เป็นแนวคิดเกี่ยวกับการประมาณค่าพารามิเตอร์ความยากของข้อสอบและใช้สำหรับการประมาณค่าของข้อมูลในโมเดลที่ทำการศึกษา

อย่างไรก็ตามแม้แนวคิดของ Horst มีลักษณะคล้ายคลึงกับแนวคิดของโมเดล MIRT ในปัจจุบัน แต่เขายังมุ่งศึกษาไปที่การวิเคราะห์น้ำหนักองค์ประกอบ (Factor Loading) มากกว่าการมุ่งศึกษาคุณลักษณะของข้อสอบและผู้สอบ

2) Christoffersson (1975) มีแนวคิดของการวิเคราะห์องค์ประกอบใกล้เคียงกับแนวคิดของโมเดล MIRT มากกว่าแนวคิดของ Horst ในการสร้างโมเดลความน่าจะเป็นของความสัมพันธ์ระหว่างการตอบข้อสอบและเวคเตอร์ค่าพารามิเตอร์ของผู้สอบ โดยใช้โมเดลปกติสะสม (Normal Ogive Model) เพื่อประมาณค่าอำนาจจำแนก (Threshold) ซึ่งคือค่าพารามิเตอร์อำนาจจำแนกของข้อสอบในโมเดล MIRT นั่นเอง โดยค่าพารามิเตอร์อำนาจจำแนก (Threshold) จะมีการแจกแจงเป็นโค้งปกติ ซึ่งมีค่าเท่ากับสัดส่วนของการตอบข้อสอบได้ถูกต้อง โดย Christoffersson นำเสนอค่าน้ำหนักองค์ประกอบ (Factor Loading) และมีการประมาณค่าอำนาจจำแนก (Threshold) ซึ่งคล้ายคลึงกับ

การวิเคราะห์ MIRT แต่มีความแตกต่างกันที่เด่นชัด 2 ประการ คือ ประการแรกมุ่งไปที่คุณลักษณะข้อสอบที่มีลักษณะต่อเนื่องตามสมมติฐานมากกว่าที่จะมุ่งศึกษาความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง ส่วนประการที่สอง ความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องไม่ได้นำเสนอในฟังก์ชันของค่าพารามิเตอร์ของข้อสอบและเวคเตอร์ของค่าพารามิเตอร์ของผู้สอบ แต่ความน่าจะเป็นในที่นี้คือ โมเดลซึ่งแทนค่าสถิติของประชากร

3) Muthen (1978) มีการนำเสนอแนวคิดของการวิเคราะห์องค์ประกอบที่ใกล้เคียงกับ MIRT มากกว่า 2 แนวคิดแรก นั่นคือ ได้นำเสนอโมเดลเวคเตอร์ p ซึ่งเป็นสัดส่วนของคะแนนที่สังเกตได้ ในลักษณะ m มิติ (m-Dimension) ของสัดส่วนคะแนนที่ตอบข้อสอบได้ถูกต้อง สามารถเขียนสมการในรูป $p = f(\theta) + \varepsilon$ โดยกำหนดให้ θ แบ่งออกเป็น 2 ส่วน คือ (θ_1', θ_2') เมื่อ θ_1' คือเวคเตอร์ของระดับค่าอำนาจจำแนก (Thresholds) ส่วน θ_2' เป็นเวคเตอร์ของส่วนประกอบที่อยู่ใต้เมตริกซ์เชิงทแยงของความสัมพันธ์แบบเตตระคลอริคของประชากร

จากการศึกษาแนวคิดของ Christoffersson (1975) และ Muthen (1978) พบว่ามีความใกล้เคียงกับแนวคิดของ MIRT ในปัจจุบัน แต่ขาดในส่วนของสูตรที่เกี่ยวข้องกับเงื่อนไขของความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องกับฟังก์ชันที่เกี่ยวข้องกับตำแหน่งของผู้สอบในมิติ θ

4) McDonald (1967) มีแนวคิดของระเบียบวิธีการวิเคราะห์องค์ประกอบแบบไม่เป็นเส้นตรง (Nonlinear Factor Analysis) ซึ่งถือได้ว่าเป็นแนวคิดที่ใกล้เคียงกับแนวคิดของ MIRT มากที่สุด เนื่องจากประสบกับปัญหาการวิเคราะห์ตัวแปรที่ให้คะแนนแบบ 2 ค่า คือ 0 และ 1 จึงแก้ปัญหาโดยการระบุองค์ประกอบความยากในการวิเคราะห์ข้อมูลแบบ 2 ค่า เพื่อจัดกระทำกับข้อมูลได้ง่ายขึ้น ถ้าในกรณีของข้อมูลที่สังเกตได้ สามารถจัดกระทำให้เป็นแบบไม่ใช่เชิงเส้นตรงได้ และ McDonald ได้ให้แนวคิดที่สำคัญของความเป็นอิสระระหว่างข้อสอบ (Local Independence) ซึ่งเป็นพื้นฐานของการวิเคราะห์ข้อสอบ โดยนำเสนอสมการพื้นฐานของความเป็นอิสระ นอกจากนี้ McDonald นำเสนอพื้นฐานของการตอบสนองข้อสอบ คือ การถดถอยของการตอบสนองข้อสอบในลักษณะมิติของลักษณะความสามารถและได้ระบุถึงความสัมพันธ์ของฟังก์ชันการถดถอย ซึ่งอยู่ภายใต้เงื่อนไขของความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง

เมื่อพิจารณาแนวคิดของ McDonald และ MIRT จะเห็นประเด็นของความแตกต่าง คือ การใช้โมเดลที่ตรวจให้คะแนนมากกว่า 2 ค่า (Polynomial Model) สามารถนำมาจัดกระทำให้อยู่ในรูปของตัวแปรที่มีการตรวจให้คะแนน 2 ค่า อยู่ในช่วงของ 0 และ 1 อย่างไรก็ตาม McDonald มุ่งไปที่การประมาณค่าองค์ประกอบมากกว่าที่จะมุ่งประเด็นของคุณลักษณะตัวแปรหรือปฏิสัมพันธ์ของตัวแปร ต่อมาในปี ค.ศ. 1985 McDonald ได้ทำให้เห็นความสัมพันธ์ของการวิเคราะห์องค์ประกอบกับ MIRT ชัดเจนมากขึ้น โดยบ่งชี้ว่า การวิเคราะห์องค์ประกอบถือเป็นกรณีเฉพาะของ IRT โดยมองว่าในการวิเคราะห์องค์ประกอบร่วมซึ่งเป็นกรณีเฉพาะของทฤษฎีการวิเคราะห์คุณลักษะแฝงตั้งอยู่บนพื้นฐานของความเป็นอิสระ (Local Independence)

5) Bock & Aitkin (1981) ได้นำเสนอแนวคิดที่คล้ายกันระหว่างการวิเคราะห์องค์ประกอบกับ IRT และผลที่ได้จากการวิเคราะห์ MIRT โดยระบุโมเดลปกติสะสม (Normal Ogive Model)

สำหรับการวัดคุณลักษณะหลายมิติ รวมทั้งคุณลักษณะของข้อสอบในรูปแบบของการวิเคราะห์องค์ประกอบ และ IRT ซึ่งส่วนใหญ่จะนำเสนอค่าความยากและอำนาจจำแนกของข้อสอบ เพื่อเป็นตัวแทนของค่าจุดตัดแกน (Intercept) และความชัน (Slopes) แต่ยังขาดในส่วนของโมเดลเต็มรูปแบบของ MIRT ใช้ตีความหมายค่าพารามิเตอร์ข้อสอบ นั่นคือ ปฏิสัมพันธ์ระหว่างผู้สอบและข้อสอบ ซึ่งการตีความหมายยังเป็นลักษณะของการวิเคราะห์องค์ประกอบ นั่นคือ พิจารณาในลักษณะขององค์ประกอบ ซึ่ง Bock & Aitkin ได้แผ่ขยายแนวคิดในลักษณะ 2 มิติ ของโมเดลปกติสะสมแบบ 2 พารามิเตอร์ ไปสู่ชนิดของการประมาณค่าพารามิเตอร์แบบ MIRT

อย่างไรก็ตามแม้ว่าโมเดลที่นำเสนอโดย Bock & Aitkin (1981) จะมีประโยชน์สำหรับการแผ่ขยายแนวคิดของโมเดล MIRT แต่จุดมุ่งหมายหลักยังคงมุ่งไปที่การวิเคราะห์องค์ประกอบเป็นสำคัญ ซึ่งคล้ายคลึงกับการศึกษาของ Bock, Gibbons, & Muraki (1988) ที่เน้นไปที่การระบุจำนวนองค์ประกอบมากกว่าที่จะพิจารณาปฏิสัมพันธ์ของผู้สอบและข้อสอบ

(2) ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory; IRT)

เมื่อพิจารณาแนวคิดของ IRT และการวิเคราะห์องค์ประกอบจะเห็นได้ว่ามีจุดมุ่งหมายในการวิเคราะห์ที่แตกต่างกัน โดยการวิเคราะห์องค์ประกอบจะมุ่งไปที่การกำหนดองค์ประกอบให้น้อยที่สุดที่ถอดแบบข้อมูลในรูปของเมตริกซ์การตอบข้อสอบ ส่วนโมเดล IRT จะศึกษาความสัมพันธ์ระหว่างผู้สอบและข้อสอบ โดย Lord (1980) ได้กล่าวว่า

> *"จุดมุ่งหมายของ IRT คือ การอธิบายลักษณะข้อสอบจากการพิจารณาค่าพารามิเตอร์ของข้อสอบและอธิบายลักษณะของผู้สอบจากการพิจารณาค่าพารามิเตอร์ความสามารถของผู้สอบ เพื่อสามารถทำนายโอกาสในการตอบข้อสอบถูกของแต่ละบุคคลในข้อสอบแต่ละข้อได้ แม้ว่าผู้สอบไม่เคยทำข้อสอบข้อนั้นมาก่อนก็ตาม ซึ่งจะพิจารณาจากลักษณะที่คล้ายคลึงกันทั้งผู้สอบและข้อสอบ"*

จากการศึกษาแนวคิดของ IRT ในช่วงแรกยังอยู่บนข้อตกลงเบื้องต้นของการอธิบายค่าพารามิเตอร์ของผู้สอบในมิติความสามารถเพียงมิติเดียว (Lord & Novick, 1968; Rasch, 1960) ซึ่งง่ายต่อการฝ่าฝืนข้อตกลงเบื้องต้นและส่งผลต่อการประมาณค่าพารามิเตอร์ไม่ถูกต้อง (Camilli, Wang & Fesq, 1995) ดังนั้นการศึกษาในช่วงแรกจึงพิจารณาลักษณะการวัดแบบพหุมิติในมุมมองของ IRT ดังที่จะนำเสนอต่อไปนี้

Rasch (1960) ในช่วงแรก Rasch ได้ศึกษาเฉพาะโมเดล UIRT ต่อมาในปี ค.ศ. 1962 ได้นำเสนอโมเดลที่ยอมรับโดยทั่วไปรวมถึงความเป็นไปได้ของความสามารถของผู้สอบ โดยนำเสนอในรูปของเวคเตอร์มากกว่าค่าสเกลลา ซึ่งสามารถแสดงได้ดังสมการที่ (2.1)

$$P\left(x\middle|\theta_j,\sigma_i\right)=\frac{1}{\gamma\left(\theta_j,\sigma_i\right)}\exp\left[\phi(x)'\theta_j+\psi(x)'\sigma_i+\theta_j'\chi(x)\sigma_i+\rho(x)\right]..(2.1)$$

เมื่อ ϕ,ψ,χ และ ρ แทน ฟังก์ชันของคะแนนในแต่ละข้อ

X แทน ค่าสถิติที่ได้จากการสังเกตได้จากค่าพารามิเตอร์ของผู้สอบและแบบสอบ

σ แทน เวคเตอร์ค่าพารามิเตอร์ของข้อสอบ

$\gamma(\ )$ แทน ฟังก์ชัน Normalizing ที่จะสามารถรับประกันได้ว่าค่าที่อยู่ในฟังก์ชันของ IRT อยู่ในช่วง 0 และ 1

จากการพิจารณาโมเดลข้างต้นจะเห็นว่ามีการระบุระดับความสามารถของผู้สอบในรูปของเวคเตอร์ค่าพารามิเตอร์ โดยการประมาณค่าพารามิเตอร์ของผู้สอบและข้อสอบมีความเป็นอิสระจากกัน เมื่อพิจารณาส่วนประกอบอื่นๆ เช่น ϕ,ψ,χ ซึ่งเป็นเวคเตอร์ของการให้คะแนนจะต้องทราบข้อมูลในส่วนนี้ก่อนทำการวิเคราะห์ด้วยโมเดลราสช์ (Rasch Model) ไม่สามารถประมาณค่าได้จากข้อมูลในการตอบของผู้สอบ สำหรับในกรณีของข้อมูลที่มีลักษณะเป็นเอกมิติ การให้คะแนนจะมี 2 ค่า คือ 0 หรือ 1 โดยไม่ได้คำนึงถึงลักษณะข้อสอบ ดังนั้นการพิจารณาเฉพาะจำนวนที่ตอบข้อสอบได้ถูกต้อง ถือเป็นค่าสถิติที่เพียงพอสำหรับการพิจารณาค่า θ

ถ้าหากมีการตอบสนองข้อสอบในมิติที่ต่างกัน 2 ด้าน โดยคะแนนการตอบข้อสอบในแต่ละข้อของมิติที่ 1 และ 2 คือ μ_1 และ μ_2 ดังนั้นค่าสถิติที่เพียงพอสำหรับประมาณค่า θ_1 และ θ_2 คือ $n\mu_1$ และ $n\mu_2$ เมื่อ n คือ จำนวนข้อที่ตอบข้อสอบได้ถูกต้อง อย่างไรก็ตามเนื่องจาก μ_1 และ μ_2 เป็นค่าคงที่ข้ามข้อสอบ ดังนั้นคะแนนที่ได้แต่ละมิติเป็นฟังก์ชันของ n เท่านั้น ดังนั้นค่า θ_1 และ θ_2 ที่ประมาณได้ จึงมีความสัมพันธ์กันอย่างสมบูรณ์ นั่นคือโมเดลที่ได้ยังคงเป็นโมเดลแบบเอกมิตินั่นเอง จากปัญหาดังกล่าวนี้ มีนักวิจัยหลายท่าน พยายามที่จะแบ่งคะแนนให้มากกว่า 2 ค่า เช่น จัดกระทำกับข้อสอบจากลักษณะ ***"Dichotomous Items"*** แปลงให้เป็น ***"Single Polytomous Item"*** Rackase (1972) เพื่อที่จะนำไปสู่การวิเคราะห์แบบ ***"Polytomous Items"*** (Kelderman, 1994) ดังนั้นฟังก์ชันการให้คะแนนจึงมีความซับซ้อนมากขึ้น หรือกำหนดให้ฟังก์ชันการให้คะแนนแตกต่างในแต่ละข้อก่อนที่จะวิเคราะห์ด้วยโมเดลราสช์ เพื่อทำให้การวิเคราะห์มีความเป็นเหตุเป็นผลมากขึ้นสำหรับการวิเคราะห์คุณลักษณะของข้อสอบ (Glass, 1992) จะเห็นได้ว่าไม่นิยมนำโมเดลราสช์มาใช้ในการวิเคราะห์แบบพหุมิติ เนื่องจากมีความซับซ้อนของกระบวนการและความถูกต้องเกี่ยวกับฟังก์ชันการให้คะแนนที่มีความเฉพาะเจาะจง

ต่อมา Lord & Novick (1968) ได้เสนอโมเดลพื้นฐานของ MIRT ที่บ่งชี้ถึงนิยามของมิติคุณลักษณะแฝงได้อย่างสมบูรณ์ และข้อตกลงเบื้องต้นของความเป็นอิสระ ซึ่งความเป็นอิสระในที่นี้หมายถึงคุณลักษณะของกลุ่มผู้สอบต่างๆที่มีค่าเหมือนกันของมิติความสามารถ $\theta_1,\theta_2,...,\theta_k$ เป็นการกระจายของคะแนนผลการตอบข้อนั้นของผู้สอบแต่ละคนไม่สัมพันธ์กัน เมื่อ k คือจำนวนของมิติ ส่วน θ ในที่นี้เป็นเวคเตอร์มิติคุณลักษณะที่สมบูรณ์ นอกจากนี้ Lord & Novick (1968) แสดงให้เห็นถึงความสัมพันธ์ระหว่างโมเดลปกติสะสมตามทฤษฎีตอบสนองข้อสอบแบบเอกมิติและโมเดลองค์ประกอบร่วมอย่างชัดเจนมากยิ่งขึ้น

นอกจากนี้ Samejima (1974) ได้เสนอสูตรที่ใช้โดยทั่วไป โดยมีข้อตกลงเบื้องต้นว่าการตอบที่มีลักษณะการตรวจให้คะแนน 2 ค่าหรือมากกว่า 2 ค่า เป็นผลมาจากการแบ่งเป็นส่วนย่อยๆ จากการตอบของตัวแปรต่อเนื่อง ซึ่ง Samejima (1974) ได้พัฒนาโมเดล MIRT สำหรับข้อสอบซึ่งมีลักษณะเป็นการตอบแบบต่อเนื่อง แสดงได้ดังสมการที่ (2.2)

$$P_{Z_i}(\theta) = 2\pi^{-1/2} \int_{-\alpha}^{a_i(\theta - b_i)} \exp\left(-\frac{u^2}{2}\right) du \quad \text{.....(2.2)}$$

แม้ว่า โมเดลของ Samejima (1974) เป็นโมเดลแรกที่ทำให้แนวความคิดเกี่ยวกับโมเดล MIRT มีความชัดเจน แต่จากการศึกษาของ Bejar (1977) พบว่าโมเดลดังกล่าวไม่สามารถนำไปประยุกต์ใช้ได้จริงในทางปฏิบัติ ซึ่งสาเหตุอาจเกิดจากการตอบข้อสอบ มีการพิจารณาเป็นตัวแปรต่อเนื่องซึ่งไม่สอดคล้องกับบริบทของแบบสอบทั้งทางด้านการศึกษาและจิตวิทยา แต่อย่างไรก็ตามถ้ามีการประเมินเกี่ยวกับทักษะการปฏิบัติ โมเดลดังกล่าวก็จะเหมาะในการประยุกต์ใช้ได้มากขึ้น (cited in Reskase, 1997)

(3) เปรียบเทียบวิธีการของการวิเคราะห์องค์ประกอบ และ MIRT (Comparison of the Factor Analysis and MIRT Approach)

จากการศึกษาถึงวิธีการวิเคราะห์องค์ประกอบและ MIRT ของ Bock & Aitkin (1981), Samejima (1974) & McDonald (1967) พบว่า สูตรทางสถิติของทั้งสองวิธีจะมีลักษณะเป็นเมตริกซ์ของการตอบข้อสอบที่ตรวจให้คะแนนแบบ 2 ค่า ซึ่งความเป็นจริง โปรแกรมคอมพิวเตอร์ที่พัฒนาโดย Bock et. al ในปี ค.ศ. 1988 สามารถนำมาประยุกต์ใช้ทั้งการวิเคราะห์องค์ประกอบและ MIRT ซึ่งต่อมา McDonald (1989) พัฒนาโปรแกรมสำหรับวิเคราะห์ตามโมเดล MIRT รวมถึงการวิเคราะห์องค์ประกอบเชิงยืนยัน เมื่อพิจารณาถึงแนวความคิดของทั้ง Factor Analysis และ MIRT พบว่ามีทั้งส่วนที่คล้ายคลึงกันและส่วนที่ต่างกัน ดังนั้นผู้วิจัยจึงสรุปความแตกต่างของทั้ง 2 วิธี ดังแสดงในตาราง

ตารางที่ 2.1 สรุปความแตกต่างของการวิเคราะห์องค์ประกอบ (Factor Analysis) และ MIRT

Methodology	Factor Analysis	MIRT
1. Focus	มุ่งไปที่เทคนิคการสกัดข้อมูลและสังเคราะห์ข้อมูลให้มีขนาดเล็กลง (Data Reduction Technique) เพื่อให้ได้มิติที่น้อยที่สุดของข้อมูลและง่ายต่อความเข้าใจสำหรับการบ่งชี้คุณลักษณะต่าง ๆ	ไม่ได้มุ่งไปที่เทคนิคการสกัดข้อมูลและสังเคราะห์ข้อมูลให้มีขนาดเล็กลงแต่มุ่งไปที่ปฏิสัมพันธ์ระหว่างผู้สอบและข้อสอบ
2. Input Variable	สนใจ Correlation หรือ Variance/Covariance Matrices มากกว่าค่าเฉลี่ยและความสัมพันธ์ระหว่างตัวแปร ซึ่ง $\bar{x}$ และ S.D. มีผลน้อยมากหรืออาจจะไม่มีผลต่อผลการวิเคราะห์ที่ได้	ค่า $\bar{x}$ และ S.D. ของการตอบข้อสอบแทนด้วยค่าพารามิเตอร์ความยากและอำนาจจำแนกของข้อสอบ ซึ่งการประมาณค่าและการตีความหมายของค่าพารามิเตอร์เหล่านี้ถือเป็นส่วนประกอบสำคัญในการวิเคราะห์
3. Data Analysis	เมตริกซ์ที่ใช้เป็นข้อมูลในการวิเคราะห์ คือ Z-Score Matrix นั่นคือ เป็นเมตริกซ์ความสัมพันธ์ (Correlation Matrices) ที่มีการแปลงคะแนนแบบเส้นตรงของคะแนนที่สังเกตได้ไปสู่ Z-Score ส่วนค่าความแปรปรวนร่วม (Covariance Matrices) คือ เมตริกซ์ของการแปลงคะแนนเชิงเส้นตรงของจำนวนคะแนนที่ตอบข้อสอบได้ถูกต้องโดยมีค่าเฉลี่ยเท่ากับ 0	เป็นเมตริกซ์ของฟังก์ชันทางคณิตศาสตร์ในการตอบสนองข้อสอบ ที่จะระบุถึงคุณลักษณะในมิติของ θ

Methodology	Factor Analysis	MIRT
4. Goodness of Fit Under Hypothesis Model	โมเดลสมมติฐานมุ่งไปที่เมตริกซ์ความแปรปรวนและเมตริกซ์ความแปรปรวนร่วมโดยภาพรวมของกลุ่มผู้สอบ มากกว่าตัวแปรใดตัวแปรหนึ่งหรือการเลือกเฉพาะกลุ่มย่อยใดกลุ่มย่อยหนึ่ง มุ่งไปที่การวัดโดยภาพรวมมากกว่าเงื่อนไขการวัดความสอดคล้องของข้อมูล	โมเดลที่แสดงปฏิสัมพันธ์ระหว่างผู้สอบและข้อสอบ มีเป้าหมายของการระบุถึงความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องแม่นยำของแต่ละบุคคลในแต่ละมิติของ θ โดยให้ความสำคัญกับข้อสอบที่ไม่สอดคล้องกับโมเดลหรือไม่สอดคล้องกับการทำนายความน่าเป็นโดยเฉพาะในช่วงพิสัยของความสามารถ ซึ่งเงื่อนไขของการวัดที่สอดคล้องกับแบบสอบชุดหนึ่งๆ จะต้องอยู่ภายใต้เงื่อนไขความสอดคล้องของผู้สอบ (Liou & Chang, 1992) และความเหมาะสมของการวัด (Drasgow, Levine & McLaughlin, 1991)
5. Analysis & Application	แม้ว่าวิธีการวิเคราะห์จะมุ่งไปที่การหมุนแกนตัวประกอบ (Procrustes Rotation) และสัมประสิทธิ์ความสอดคล้อง เพื่อที่จะแสวงหาองค์ประกอบร่วมกันแต่ยังมีลักษณะที่เน้นย้ำน้อยกว่า MIRT	วิธีการวิเคราะห์จะมีการศึกษาข้ามแบบสอบและกลุ่มผู้สอบ จุดมุ่งหมายของการวิเคราะห์เพื่อให้อยู่ในสเกลร่วมกัน โดยมีประโยชน์ในการจัดเก็บเป็นคลังข้อสอบ หรือการจัดชุดของแบบสอบเพื่อวัดโครงสร้างหรือคุณลักษณะที่มุ่งวัด (Fixed-Form Test Construction) หรือ การสร้างแบบสอบให้เหมาะสมกับความสามารถของผู้สอบ (Adaptive Testing)

1.1.2 พัฒนาการของโมเดล MIRT

ในช่วงปลายปี ค.ศ. 1971-1979 ถึงช่วงต้นปี ค.ศ. 1981-1989 พบว่าเริ่มมีการพัฒนาแนวคิดเกี่ยวกับ MIRT มากขึ้นเพื่อให้เกิดผลในทางปฏิบัติ ดังจะเห็นได้จากการศึกษาเกี่ยวกับโมเดลราสช์แบบพหุมิติ (Multidimensional Rasch Model) ที่ทำการศึกษาโดย Reckase (1972) ต่อมาก็มีการขยายแนวคิดของโมเดล MIRT ในการศึกษาปฏิสัมพันธ์ระหว่างผู้สอบและข้อสอบ (Mulaik, 1972; Sympsom, 1978; Whitely, 1980) โดยมีรายละเอียดดังนี้

โมเดลของ Mulaik ที่นำเสนอในปี ค.ศ. 1972 สามารถแสดงได้ดังสมการ

$$P\left(x_{ij}\middle|\theta_j,\sigma_i\right)=\frac{\sum_{k=1}^{m}\exp(\theta_{jk}+\sigma_{ik})x_{ij}}{1+\sum_{k=1}^{m}\exp(\theta_{jk}+\sigma_{ik})} \quad \text{.....(2.3)}$$

เมื่อ x_{ij} = 0, 1

จากโมเดลที่แสดงในสมการที่ (2.3) ถือว่าเป็นโมเดลที่มีคุณสมบัติน่าสนใจ นั่นคือ มีการกำหนดค่าสมการอย่างชัดเจน โดยค่าความน่าจะเป็นในการตอบข้อสอบถูก เพิ่มขึ้นเช่นเดียวกับจำนวนของ

มิติเพิ่มขึ้น ซึ่งถ้าสมการทั้งหมด เท่ากับ 0 ความน่าจะเป็นในการตอบข้อสอบได้ถูกมีค่าเท่ากับ m/(m+1) ถ้าค่าพารามิเตอร์สามารถตีความหมายเป็นค่าคงที่ แสดงให้เห็นว่าถ้าจำนวนมิติมีการเปลี่ยนแปลง คุณสมบัติดังกล่าวสามารถส่งผลต่อค่าพารามิเตอร์ข้อสอบนั่นเอง

ต่อมา Sympsom (1978) & Whitely (1980) ได้เสนอโมเดลที่มีคุณสมบัติในลักษณะเดียวกัน แต่มีความสัมพันธ์ในลักษณะที่ตรงกันข้ามกัน โดยกำหนดค่าของส่วนประกอบในสมการ ซึ่งถ้าค่าความน่าจะเป็นในการตอบข้อสอบถูกลดลง จำนวนของมิติจะเพิ่มขึ้น สามารถแสดงได้ดังสมการ

$$P\left(x_{ij}=1\middle|\theta_j,a_{ij},b_i,c_i\right)=c_i+\left(1-c_i\right)\prod_{k=1}^{m}\frac{\exp\left[a_{ik}\left(\theta_{jk}-b_{ik}\right)\right]}{1+\exp\left[a_{ik}\left(\theta_{jk}-b_{ik}\right)\right]} \quad \text{.....(2.4)}$$

เมื่อ a_i เป็นเวคเตอร์ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ

b_i เป็นเวคเตอร์ค่าพารามิเตอร์ความยากของข้อสอบ

c_i เป็นเวคเตอร์โอกาสการเดาของข้อสอบ

จากโมเดลในสมการที่ (2.4) จะเห็นว่าเมื่อ สมการทั้งหมด เท่ากับ 0 ความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง มีค่าเท่ากับ $c_i+\left[\left(1-c_i\right)\left(0.5\right)^m\right]$ เช่น ในกรณีที่มิติ m เพิ่มขึ้นค่าที่ได้จะมีค่าเข้าใกล้ c_i ซึ่งโมเดลนี้มีชื่อเรียกว่า **Partially Compensatory Model** หรือ **Noncompensatory Model** เนื่องจากการเพิ่มขึ้นของ θ ใน 1 มิติ (1-Dimension) ไม่สามารถไปเพิ่มเติมในส่วนที่ขาดหายไปในมิติอื่นๆ ได้

ในปี ค.ศ. 1982 McKinley & Reckase ได้ศึกษาความแปรเปลี่ยนของโมเดลราสช์โดยทั่วไป (General Rasch Model) และรูปแบบของโมเดลโลจิสติกที่เป็นที่นิยมในปัจจุบัน ซึ่งสามารถแสดงได้ดังสมการ

$$P\left(x_{ij}=1\middle|\theta_j,a_{ij},d_i\right)=\frac{\exp\left(\sum_{k=1}^{m}a_{ik}\theta_{jk}+d_i\right)}{1+\exp\left(\sum_{k=1}^{m}a_{ik}\theta_{ik}+d_i\right)} \quad \text{.....(2.5)}$$

เมื่อ a_i เป็นเวคเตอร์ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ

d_i เป็นเวคเตอร์ค่าพารามิเตอร์ความยากของข้อสอบ

จากโมเดลดังสมการที่ (2.5) ถือเป็นโมเดลแบบพหุ (Multivariate) ที่ขยายแนวคิดมาจากโมเดลโลจิสติกแบบ 2 พารามิเตอร์ ซึ่งมีชื่อเรียกเฉพาะว่า **Compensatory Model** เพราะว่าการที่ค่า θ มีค่าลดลงในมิติใดมิติหนึ่ง สามารถไปชดเชยให้ค่า θ ในมิติอื่นๆ สูงขึ้นได้นั่นเอง ต่อมา Spray, Davey, Reckase, Ackerman, & Carlson (1990) ได้พัฒนาโมเดลที่มีนัยทั่วไป (Generalized Model) โดยเฉพาะ **Compensatory Model** และ **Partial Compensatory Model** เป็นกรณีพิเศษ แต่โมเดลดังกล่าวไม่สามารถนำมาวิเคราะห์ในข้อมูลจริงจากการตอบข้อสอบได้

กล่าวโดยสรุป โมเดล MIRT ถือเป็นแนวคิดที่แผ่ขยายมาจากโมเดล UIRT หรือเป็นกรณีเฉพาะของการวิเคราะห์องค์ประกอบหรือโมเดลสมการเชิงโครงสร้างนั่นเอง ซึ่งจะแตกต่างจากแนวคิดแบบเดิมในส่วนของระเบียบวิธีของ MIRT ที่นำมาประยุกต์และการแปลความหมายของผลที่ได้ โดยการวิเคราะห์องค์ประกอบมุ่งไปที่เทคนิคการสกัดข้อมูลและสังเคราะห์ข้อมูลให้มีขนาดเล็กลง (Data Reduction Technique) เพื่อให้ได้มิติที่น้อยที่สุดของข้อมูลและง่ายต่อความเข้าใจสำหรับการบ่งชี้คุณลักษณะต่าง ๆ ส่วนโมเดล MIRT ไม่ได้มุ่งไปที่เทคนิคการสกัดข้อมูลและสังเคราะห์ข้อมูลให้มีขนาดเล็กลง แต่มุ่งไปที่ปฏิสัมพันธ์ระหว่างผู้สอบและข้อสอบ ซึ่งพบว่าในช่วงปลายปี ค.ศ. 1971-1979 ถึงช่วงต้นปี ค.ศ. 1981-1989 เริ่มมีการพัฒนาแนวคิดเกี่ยวกับ MIRT มากขึ้นเพื่อให้เกิดผลในทางปฏิบัติ ซึ่งจากพัฒนาการ MIRT ที่กล่าวมาข้างต้นจะเห็นได้ว่า สิ่งสำคัญที่สุดในการที่จะบอกว่าโมเดลใดเหมาะสมที่สุด คือกระบวนการในการประมาณค่าพารามิเตอร์ จากการศึกษาของ Mulaik (1972) และ Spray et al. (1990) มีข้อสังเกตว่า โมเดลขาดความชัดเจนในกระบวนการประมาณค่าพารามิเตอร์จึงยากต่อการพัฒนาแนวคิดต่อไป แสดงให้เห็นว่ากระบวนการประมาณค่าพารามิเตอร์ตามโมเดล MIRT ถือเป็นสิ่งสำคัญมากที่สุดและมีความท้าทายในการพัฒนา เนื่องจากการศึกษาที่ผ่านมาจนถึงปัจจุบันค่าสถิติที่สามารถสังเกตได้ยังไม่เพียงพอต่อการประมาณค่าพารามิเตอร์ของผู้สอบและข้อสอบได้อย่างเป็นอิสระ ซึ่งประสบปัญหาเช่นเดียวกับการวิเคราะห์องค์ประกอบและโมเดลสมการเชิงโครงสร้าง ในเรื่องของระเบียบวิธีในการประมาณค่า เช่นเดียวกัน

อย่างไรก็ตามจากการศึกษาความเป็นมาและพัฒนาการของโมเดล MIRT ถือว่าเป็นการจุดประกายแนวคิดสำคัญที่นำไปสู่การพัฒนาวิธีการศึกษาแบบพหุมิติ และนำไปสู่การนำไปประยุกต์ใช้มากขึ้น ซึ่ง Embretson & Reise (2000) ได้รวบรวมแนวคิดเกี่ยวกับโมเดล MIRT ในหนังสือ “Item Response Theory for Psychologists” โดยแบ่งประเภทของโมเดล MIRT ออกเป็น 2 ประเภท ประเภทแรกคือ Exploratory MIRT Model ประกอบด้วย 2 ประเด็นย่อย ดังนี้ 1) Multidimensional Logistic Model และ 2) Multidimensional Normal Ogive Model ส่วนโมเดลประเภทที่สองคือ Confirmatory MIRT ประกอบด้วย 4 ประเด็นย่อยดังนี้ 1) Models for Noncompensatory Dimensional, 2) Models for Learning and Change, 3) Models With Specified Trait Level Structures และ 4) Models for Distinct Classes of Persons โดยจะกล่าวถึงในรายละเอียดต่อไป

1.2 ประเภทของโมเดล MIRT

โมเดล MIRT ประกอบด้วยค่าพารามิเตอร์ของผู้สอบตั้งแต่ 2 ค่าขึ้นไป โดยการศึกษาในหลายมิติ จะทำให้โมเดลมีความสอดคล้องกับข้อมูลมากขึ้น สำหรับผู้สอบที่มีความแตกต่างกันอย่างเป็นระบบ เนื่องจากความยากของชุดข้อสอบ โมเดลหลายโมเดลที่เป็นแบบพหุมิติ ค่าอำนาจจำแนกของข้อสอบจะส่งผลกระทบต่อมิติในแต่ละชุดข้อสอบ โดยมีสัญลักษณ์ที่แสดงในโมเดล ดังนี้

X_{is} = ผลการตอบของผู้สอบที่ s ในข้อที่ i

θ_{sm} = ระดับความสามารถ (Level Trait) สำหรับบุคคลที่ s ในมิติ (Dimension) ที่ m

β_i = ค่าความยากของข้อที่ i

δ_i = Easiness Intercept สำหรับข้อที่ i

α_{im} = ค่าอำนาจจำแนกของข้อที่ i ในมิติ (Dimension) ที่ m

γ_i = ค่าโอกาสการเดาข้อสอบของข้อที่ i

การนำเสนอโมเดลแบบพหุมิติ Embretson & Reise (2000) ได้เสนอโมเดล MIRT เป็น 2 ประเภท คือ โมเดลการตอบสนองข้อสอบแบบพหุมิติเชิงสำรวจ (Exploratory MIRT Model) และโมเดลการตอบสนองข้อสอบแบบพหุมิติเชิงยืนยัน (Confirmatory MIRT Model) ซึ่งมีลักษณะคล้ายคลึงกับการวิเคราะห์องค์ประกอบ โดย Exploratory MIRT Model เป็นโมเดลที่เกี่ยวข้องกับการประมาณค่าพารามิเตอร์ของข้อสอบและผู้สอบที่มากกว่า 1 มิติ เพื่อให้โมเดลมีความสอดคล้องกับข้อมูลโดยไม่ได้กำหนดจำนวนขององค์ประกอบไว้ล่วงหน้า ซึ่งตรงข้ามกับ Confirmatory MIRT Model เป็นโมเดลที่เกี่ยวข้องกับการประมาณค่าพารามิเตอร์ในมิติที่มีความเฉพาะเจาะจง ซึ่งคล้ายคลึงกับการวิเคราะห์องค์ประกอบ โดยการวิเคราะห์จะเกี่ยวข้องกับการระบุความสัมพันธ์ของข้อสอบไปยังมิติต่างๆ โดยมีรายละเอียดในแต่ละโมเดลดังต่อไปนี้

(1) โมเดลการตอบสนองข้อสอบแบบพหุมิติเชิงสำรวจ (Exploratory MIRT Model)

ในการวิเคราะห์องค์ประกอบซึ่งมีลักษณะคล้ายกับโมเดล MIRT ดังที่กล่าวมาแล้วข้างต้น ความจริงแล้วเมื่อพิจารณาตามข้อตกลงเบื้องต้น สามารถพิสูจน์ได้ว่าเป็นโมเดลเดียวกัน (Takane & de Leeuw, 1988) ซึ่ง McDonald (1967) ได้เสนอโมเดลการวิเคราะห์องค์ประกอบแบบไม่ใช่เส้นตรง โดยสร้างจากการรวมพื้นฐานของการวิเคราะห์องค์ประกอบ CTT และ IRT เข้าด้วยกัน ซึ่งปัจจุบันจากการศึกษาของ Reckase (1997) ได้ระบุว่าขอบเขตของ Exploratory Factor Analysis จะนิยมใช้กับข้อมูลแบบแบ่ง 2 คล้ายกับโมเดลของ IRT จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับ Exploratory MIRT Model พบว่ามีความสัมพันธ์กันในระดับสูงกับการวิเคราะห์องค์ประกอบ โดยศักยภาพของผู้สอบพิจารณาจากการตอบข้อสอบ คือ การให้น้ำหนักรวมกันภายใต้มิติของคุณลักษณะที่ต้องการศึกษา

ปัจจุบันนี้โมเดลของ MIRT ส่วนใหญ่จะช่วยในการพิจารณาด้านโครงสร้างของการพัฒนาแบบสอบ นั่นคือการพิจารณาโครงสร้างที่ต้องการวัดในข้อสอบจำเป็นที่จะต้องอาศัยหลักการของโมเดล IRT เพื่อให้สอดคล้องกับข้อมูลโดยจำนวนของคุณลักษณะที่ต้องการวัดจะต้องมีความสอดคล้องกับข้อมูลเชิงประจักษ์ และเป็นส่วนย่อยของมิติต่างๆ เป็นการสร้างตามหลักการหรือทฤษฎีในชุดข้อสอบ ผลที่ได้จากการประมาณค่าพารามิเตอร์ของข้อสอบจะเป็นประโยชน์ในการเลือกข้อสอบเพื่อให้เหมาะสมกับคุณลักษณะแฝง นั่นคือ แม้ว่าลักษณะที่วัดจะมีความเกี่ยวข้องกันบางส่วนก็ตามแต่สามารถเลือกศึกษาเพียงคุณลักษณะใดลักษณะหนึ่งได้

อย่างไรก็ตามเราสามารถประมาณค่าพารามิเตอร์ของผู้สอบที่มีหลายคุณลักษณะแฝงไปพร้อมๆ กันได้ในครั้งเดียวโดยใช้โมเดล MIRT ซึ่งข้อมูลที่ได้จากการตอบของผู้สอบถือว่าเป็นข้อมูลที่สำคัญในการวัดตั้งแต่ 2 คุณลักษณะขึ้นไป โดยในที่นี้ได้จำแนก Exploratory MIRT Model ออกเป็นประเภทย่อย ได้ 2 ประเภท คือ 1) Multidimensional Logistic Model และ 2) Multidimensional Normal Ogive Model ซึ่งมีรายละเอียดดังนี้

(1.1) โมเดลโลจิสติกแบบพหุมิติ (Multidimensional Logistic Model)

เป็นที่ทราบกันดีว่าโมเดลโลจิสติกตามทฤษฎี UIRT จะประกอบด้วยโมเดลโลจิสติก ชนิด 1, 2 และ 3 พารามิเตอร์ โมเดล MIRT ก็เช่นเดียวกัน จะประกอบไปด้วยโมเดลโลจิสติก ชนิด 1, 2 และ 3 พารามิเตอร์ แต่จะต่างกันในจำนวนมิติที่ศึกษา โดยมีรายละเอียดดังนี้

(1.1.1) โมเดลโลจิสติกแบบพหุมิติชนิด 1 พารามิเตอร์ (Multidimensional Rasch Model)

โมเดลพหุมิติชนิด 1 พารามิเตอร์ ได้มีการนำเสนอโดย McKinley & Reckase (1982) เป็นโมเดลโลจิสติกที่สามารถแสดงได้ดังสมการ

$$P\left(x_{is}=1|\underline{\theta}_s,\delta_i\right)=\frac{\exp\left(\sum_m \theta_{sm}+\delta_i\right)}{1+\exp\left(\sum_m \theta_{sm}+\delta_i\right)} \quad \text{.....(2.6)}$$

จากสมการที่ (2.6) เมื่อเปรียบเทียบกับโมเดลราสช์แบบเอกมิติ (Unidimensional Rasch Model) ซึ่งเป็นโมเดลอย่างง่าย จะพบว่าระดับความสามารถในลักษณะ 1 มิติ ของบุคคลที่ s เขียนแทนด้วย θ_s จะถูกแทนด้วยการถ่วงน้ำหนักด้วยการรวมหลายลักษณะที่มุ่งวัด ซึ่ง Reckase (1978) ได้ชี้ว่าเมื่อมีการศึกษาในหลายมิติต้องมีการถ่วงน้ำหนักของแต่ละข้อ ซึ่งโมเดลราสช์แบบเอกมิติ จะมีความสอดคล้องกับข้อมูล นอกจากนี้ยังพบว่าไม่มีแบบแผนที่แน่นอนในการที่จะบ่งบอกได้ถึงความแตกต่างระหว่างบุคคล ในลักษณะที่มีการศึกษาหลายมิติ ดังนั้นความแตกต่างของคุณลักษณะไม่สามารถประมาณค่าแยกออกมาและไม่มีการกำหนดโมเดลไว้ล่วงหน้า

ต่อมา Stegelmann (1983) มีความมุ่งหวังที่จะให้โมเดลราสช์แบบพหุมิติ (Multidimensional Rasch Model) สามารถประมาณค่าความแตกต่างของระดับความสามารถหรือคุณลักษณะแฝง อย่างไรก็ตาม โมเดลค่อนข้างมีข้อจำกัดในเรื่องของข้อสอบจะต้องมีค่าความยากเท่ากัน ดังนั้น โมเดลของ Stegelmann (1983) จึงไม่เหมาะสมในการนำไปประยุกต์ใช้ในทางปฏิบัติ

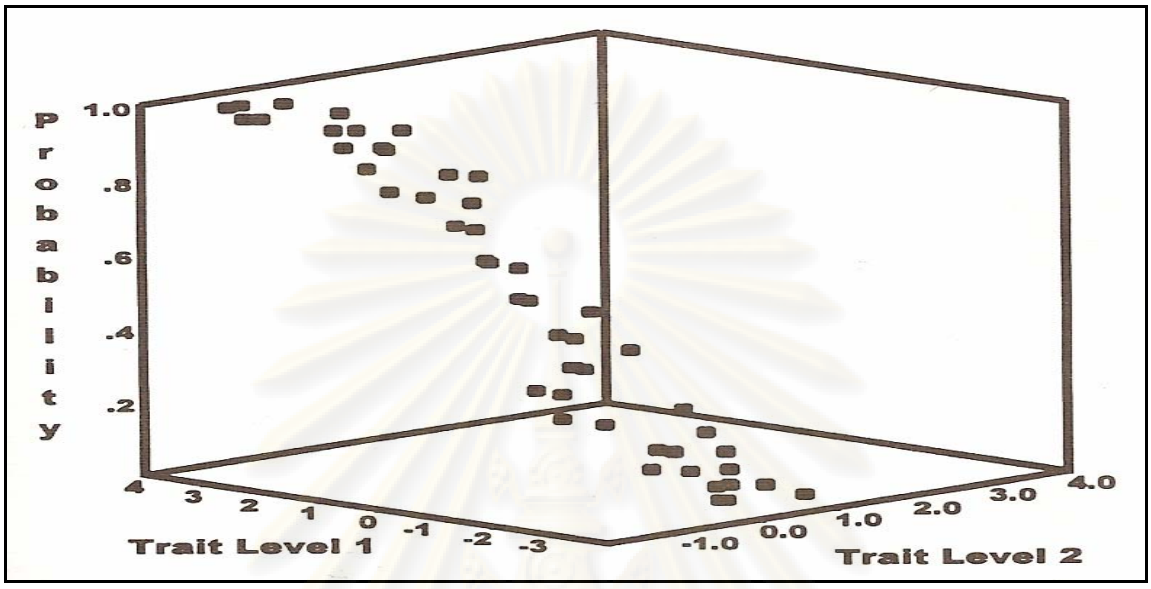
(1.1.2) โมเดลโลจิสติกแบบพหุมิติชนิด 2 พารามิเตอร์ (Multidimensional 2-Parameter Logistic Model)

สำหรับโมเดลดังกล่าวค่าอำนาจจำแนกของข้อสอบจะแตกต่างกันในแต่ละข้อของแต่ละมิติที่มุ่งศึกษา ดังแสดงในสมการ

$$P\left(x_{is}=1|\underline{\theta}_s,\delta_i,\underline{\alpha}_i\right)=\frac{\exp\left(\sum_m \alpha_{im}\theta_{sm}+\delta_i\right)}{1+\exp\left(\sum_m \alpha_{im}\theta_{sm}+\delta_i\right)} \quad \text{.....(2.7)}$$

จากสมการที่ 2.7 ศักยภาพของผู้สอบที่ทำข้อสอบแต่ละข้อจะพิจารณาจากผลการรวมของการถ่วงน้ำหนักความสามารถของคุณลักษณะมากกว่าที่จะศึกษาเพียงคุณลักษณะใดคุณลักษณะหนึ่งเท่านั้น ถ้าถ่วงน้ำหนักแล้วมีค่าพารามิเตอร์ (เช่น α_{im}; ค่าอำนาจจำแนกของ

ข้อสอบ) สูงกว่าคุณลักษณะอื่นๆ แสดงว่าคุณลักษณะนั้นมีความสำคัญ ซึ่ง Rackase (1997) ได้อธิบายว่าโมเดลนี้อาจมีข้อจำกัดไม่มาก ในเรื่องของการกำหนดค่าเฉลี่ยของค่าความสามารถลักษณะและค่าส่วนเบี่ยงเบนมาตรฐาน



แผนภาพที่ 2.1 Three-Dimensional Plot from the Multidimensional Logistic Model

จากแผนภาพที่ 2.1 เป็นการพล็อตกราฟในลักษณะ 3 มิติของความน่าจะเป็นในการตอบข้อสอบได้ถูกใน 2 คุณลักษณะแฝง ซึ่งจะเห็นว่าความแตกต่างระหว่างบุคคลของ Trait Level 1 มีความสำคัญมากต่อการประมาณค่าพารามิเตอร์ของข้อสอบมากกว่า เนื่องจากพื้นผิวของความน่าเป็นของโอกาสในการตอบข้อสอบมีการเปลี่ยนแปลงอย่างรวดเร็วส่งผลให้มีค่าอำนาจจำแนกสูงกว่า ส่วนค่าความแตกต่างระหว่างบุคคลใน Trait Level 2 ยังถือได้ว่ามีความสำคัญเช่นกัน แต่มีผลกระทบต่อความน่าเป็นของโอกาสในการตอบข้อสอบได้ถูกน้อยกว่า

(1.1.3) โมเดลโลจิสติกแบบพหุมิติชนิด 3 พารามิเตอร์ (Multidimensional 3-Parameter Logistic Model)

โมเดลโลจิสติกแบบพหุมิติ 3 พารามิเตอร์ เป็นโมเดลที่ขยายมาจากโมเดลโลจิสติก ชนิด 3 พารามิเตอร์ ซึ่งเพิ่มค่าพารามิเตอร์โอกาสในการเดาข้อสอบได้ถูก (Lower Asymptote) ดังสมการ

$$P\left(x_{is}=1|\underline{\theta}_s,\delta_i,\underline{\alpha}_i,\gamma_i\right)=\gamma_i+\left(1-\gamma_i\right)\frac{\exp\left(\sum_m \alpha_{im}\theta_{sm}+\delta_i\right)}{1+\exp\left(\sum_m \alpha_{im}\theta_{sm}+\delta_i\right)} \quad \text{.....(2.8)}$$

พื้นผิวของความน่าจะเป็นในลักษณะ 3 มิติ มีลักษณะคล้ายกับแผนภาพที่ 2.1 ยกเว้นพื้นผิวของความน่าจะเป็นไม่ได้เริ่มที่ตำแหน่งต่ำสุด จึงทำให้ความน่าจะเป็นมีค่ามากกว่า 0 เมื่อมีการรวม Trait Levels แล้ว

(1.2) โมเดลปกติสะสมแบบพหุมิติ (Multidimensional Normal Ogive Model)

เมื่อพิจารณาถึงโมเดลปกติสะสมตามทฤษฎี IRT ก็มีส่วนที่คล้ายคลึงกันระหว่าง UIRT และ MIRT ซึ่งประกอบด้วยโมเดลปกติสะสม ชนิด 1, 2 และ 3 พารามิเตอร์ แต่จะแตกต่างกันในจำนวนมิติที่ทำการศึกษา โดยมีรายละเอียดดังนี้

(1.2.1) โมเดลปกติสะสมแบบพหุมิติ ชนิด 2 พารามิเตอร์ (Multidimensional 2-Parameter Normal Ogive Model)

ในปี ค.ศ. 1988 Bock, Gibbons, & Muraki มีการสร้างโมเดลปกติสะสมแบบพหุมิติแบบ 2 พารามิเตอร์ขึ้นเป็นครั้งแรก โดยอธิบายว่า เป็นโมเดลคล้ายกับการวิเคราะห์องค์ประกอบเต็มรูปแบบ แต่จะแตกต่างกันที่การวิเคราะห์องค์ประกอบเป็นโมเดลความสัมพันธ์ (Correlation Model) แต่โมเดลปกติสะสมแบบพหุมิติ ชนิด 2 พารามิเตอร์ เป็นโมเดลการตอบข้อสอบโดยตรง ซึ่งแผ่ขยายมาจากโมเดลปกติสะสมแบบ 2 พารามิเตอร์ โดยสามารถแสดงสูตรของความสามารถของผู้สอบที่ s ของข้อสอบข้อที่ i (Z_{is}) ได้ดังสมการ

$$z_{is} = \sum_m \alpha_{im}\theta_{sm} + \delta_i \qquad \text{.....(2.9)}$$

โมเดลแบบปกติสะสมจะแสดงถึงโอกาสในการตอบข้อสอบได้ถูกต้องของผู้สอบ ซึ่งมีการแจกแจงเป็นแบบปกติ โดยสามารถแสดงได้ดังสมการ

$$P\left(x_{is}=1|\underline{\theta}_s,\delta_i,\underline{\alpha}_i\right)=\int_{-Z_{is}}^{\alpha}\frac{1}{(2\pi)^{1/2}}\exp\left(-t^2/2\right)dt \qquad \text{.....(2.10)}$$

เมื่อ λ_{im} แทน ค่าน้ำหนักองค์ประกอบ (Factor Loading) โดยที่ $\lambda_{im}=\alpha_{im}/g_i$

β_i แทน ค่ามาตรฐานความยากของข้อสอบ (Standard Item Difficulty) โดยที่

$$\beta_i=\delta_i/g_i$$

โดยกำหนดให้ $g_i=\sqrt{1+\sum_m\alpha_{im}^2}$

(1.2.2) โมเดลปกติสะสมแบบพหุมิติกับโอกาสการเดาข้อสอบ (Multidimensional Normal Ogive Model With Guessing)

Bock et. al. (1988) ได้เสนอโมเดลเกี่ยวกับโอกาสในการเดาข้อสอบ โดยเฉพาะข้อสอบแบบเลือกตอบ มีการเพิ่มค่าพารามิเตอร์โอกาสในการเดาข้อสอบ (Lower Asymtote; γ_i) ซึ่งสามารถแสดงได้ ดังสมการ

$$P\left(x_{is}=1|\underline{\theta}_s,\beta_i,\delta_i,\underline{\alpha}_i,\gamma_i\right)=\gamma_i+\left(1-\gamma_i\right)\int_{-Z_{is}}^{\alpha}\frac{1}{(2\pi)^{1/2}}\exp\left(-t^2/2\right)dt \qquad \text{.....(2.11)}$$

จากสมการที่ (2.11) จะเห็นว่ามีการเพิ่มค่าพารามิเตอร์โอกาสการเดาข้อสอบเข้ามาในสูตรของโมเดลปกติสะสม ซึ่งพบว่าค่าที่ได้มีความคงที่ข้ามกลุ่มข้อสอบหรือชุดของแบบสอบย่อย

(2) โมเดลการตอบสนองข้อสอบแบบพหุมิติเชิงยืนยัน (Confirmatory Multidimensional IRT Model)

สำหรับโมเดลการตอบสนองข้อสอบแบบพหุมิติเชิงยืนยัน พบว่าข้อสอบจะประกอบด้วยการออกแบบตามโครงสร้างและเป็นโมเดลในรูปฟังก์ชันเชิงคณิตศาสตร์ที่มีการเชื่อมโยงกับคุณลักษณะที่มีความเฉพาะเจาะจงในหลายลักษณะ ซึ่งโมเดลนี้จะมีความเหมาะสมสำหรับข้อสอบหรือชิ้นงานในลักษณะของทักษะการปฏิบัติ ที่มีการกำหนดหรือสร้างทฤษฎีไว้ล่วงหน้า ซึ่งจะได้นำเสนอตัวอย่างในลำดับต่อไป ซึ่ง Embretson & Reise (2000) ได้แบ่งโมเดลย่อย ทั้งหมด 4 โมเดล คือ 1) Models for Noncompensatory Dimensional, 2) Models for Learning and Change, 3) Models With Specified Trait Level Structures และ 4) Models for Distinct Classes of Persons โดยมีรายละเอียดดังต่อไปนี้

(2.1) โมเดลสำหรับการวัดมิติที่ไม่สามารถทดแทนได้ (Models for Noncompensatory Dimensional)

ในปี ค.ศ. 1980 Whitely ได้เสนอโมเดล Multicomponent Latent Trait Model (MLTM) เพื่อวัดกระบวนการในหลายด้าน ภายใต้ข้อสอบที่มีลักษณะของการพิสูจน์หรือแก้ปัญหา โดย MLTM จะมีการรวมโมเดลทางคณิตศาสตร์ของกระบวนการในการตอบข้อสอบร่วมกับโมเดล IRT ซึ่งศึกษาทั้งในส่วนของ Trait Level และการประมาณค่าความยากของข้อสอบ อยู่ภายใต้ข้อตกลงเบื้องต้นความถูกต้องในการประมาณค่าพารามิเตอร์ ขึ้นอยู่กับการประมาณค่าองค์ประกอบทั้งหลาย ถ้าองค์ประกอบไม่สามารถประมาณค่าได้ ข้อสอบจะไม่สามารถประมาณค่าได้เช่นกัน ดังนั้นความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องของบุคคลที่ s สำหรับข้อสอบข้อที่ i โดยรวมทุกมิติ แทนด้วย X_{isT} เกิดจากการรวมของความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องของบุคคลที่ s สำหรับข้อสอบข้อที่ i ในมิติที่ m แทนด้วย X_{ism} ซึ่งโมเดล MLTM สามารถแสดงได้ดังสมการ

$$P\left(x_{isT}=1\middle|\underline{\theta}_s,\underline{\beta}_i\right)=\prod_m \frac{\exp\left(\theta_{sm}-\beta_{im}\right)}{1+\exp\left(\theta_{sm}-\beta_{im}\right)} \quad \text{.....(2.12)}$$

เมื่อ $\underline{\theta}_s$ แทน ระดับความสามารถของผู้สอบที่ j ของ M ส่วนประกอบ (M-component)

$\underline{\beta}_s$ แทน ค่าความยากของข้อสอบข้อที่ i ของ M ส่วนประกอบ (M-component)

θ_{sm} แทน ระดับความสามารถของผู้สอบที่ j ของส่วนประกอบที่ m

β_{im} แทน ค่าความยากของข้อสอบข้อที่ i ของส่วนประกอบที่ m

เป็นที่น่าสังเกตว่าสมการทางขวามือ คือโมเดลแบบราสช์ ซึ่งเป็นความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องขององค์ประกอบที่ m ซึ่งผลคูณความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องของแต่ละองค์ประกอบคือความน่าจะเป็นทั้งหมดในการตอบข้อสอบได้ถูกต้องสำหรับข้อสอบข้อที่ i

เมื่อพิจารณา General Component Latent Trait Model (GLTM) ซึ่งเป็นแนวคิดของ Embretson (1984) เป็นโมเดลที่มีการรวมค่าความยากของข้อสอบ มีการถ่วงน้ำหนักโดยรวมองค์ประกอบต่างๆ ของสิ่งเร้าเข้าด้วยกัน แทนด้วยสัญลักษณ์ q_{ikm} โมเดลดังกล่าวแสดงได้ดังสมการ

$$P\left(x_{isT}=1\middle|\underline{\theta}_s,\underline{\beta}_i\right)=\prod_m \frac{\exp\left(\theta_{sm}-\sum_m \tau_{km}q_{ikm}\right)}{1+\exp\left(\theta_{sm}-\sum_m \tau_{km}q_{ikm}\right)} \quad \text{.....(2.13)}$$

เมื่อ τ_{ikm} แทน น้ำหนักองค์ประกอบของสิ่งเร้าที่ k ในส่วนประกอบที่ m

q_{ikm} แทน คะแนนองค์ประกอบของสิ่งเร้าที่ k ในส่วนประกอบที่ m สำหรับข้อสอบข้อที่ i

โดยทั่วไป โมเดล MLTM และ GLTM มีการกำหนดส่วนประกอบของการตอบเหมือนกับการตอบข้อสอบโดยรวมของแต่ละข้อ เพื่อใช้ในการประมาณค่าพารามิเตอร์ของส่วนประกอบการตอบข้อสอบ ซึ่งเป็นสิ่งที่สามารถสังเกตได้ เช่น การดำเนินการสอบของแบบสอบย่อย หรือชิ้นงานย่อย (Subtasks) โดยพิจารณาจากการตอบข้อสอบโดยรวมของแต่ละข้อ อย่างไรก็ตามในปัจจุบันพบว่า Marris (1995) ได้พัฒนาการประมาณค่าส่วนประกอบโดยการนำโมเดล GLTM มาประยุกต์ใช้เพื่อให้เป็นที่ยอมรับกันโดยทั่วไป โดยไม่จำเป็นต้องมีชิ้นงานย่อย การวัดโดยใช้ข้อสอบหรือใช้ชิ้นงาน ทำให้ได้โมเดลทางคณิตศาสตร์ที่มีความเหมาะสมเพื่อใช้ในการทำนายค่าพารามิเตอร์ความยากของข้อสอบหรือค่าคงที่ เพื่อใช้ในการประมาณค่าพารามิเตอร์ขององค์ประกอบ

(2.2) โมเดลสำหรับการวัดการเรียนรู้และการเปลี่ยนแปลง (Models for Learning and Change)

ในปี ค.ศ. 1991 Embretson ได้พัฒนาโมเดลที่เรียกว่า Multidimensional Rasch Model for Learning and Change (MRCML) เพื่อใช้ในการแก้ปัญหาทางด้านการวัดทางจิตวิทยาเพื่อการวัดความเปลี่ยนแปลง ซึ่งการออกแบบที่เหมาะสมสำหรับโมเดล MRCML คือ ผู้สอบจะถูกสังเกตภายใต้เงื่อนไขตั้งแต่ 2 เงื่อนไขขึ้นไป เช่น การทดสอบ Pre-Test และ Post-Test เป็นต้น ในการพิจารณาทักษะการปฏิบัติ โดยมีการแยกเงื่อนไขออกจากกัน และนำเอากระบวนการทางโครงสร้างของ Wiener มาใช้ในการระบุมิติความสามารถ แทนด้วยสัญลักษณ์ θ_m โดยมีเงื่อนไขเฉพาะ c_k ในที่นี้ผู้วิจัยได้นำเสนอการออกแบบ สำหรับ 4 เงื่อนไขของการวัด ตามโครงสร้างที่กำหนดไว้ สามารถแสดงได้ดังแผนภาพ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

$$\Lambda_{Kxm} = \begin{array}{c|cccc} & \underline{\theta}_1 & \underline{\theta}_2 & \underline{\theta}_3 & \underline{\theta}_4 \\ c_1 & 1 & 0 & 0 & 0 \\ c_2 & 1 & 1 & 0 & 0 \\ c_3 & 1 & 1 & 1 & 0 \\ c_4 & 1 & 1 & 1 & 1 \end{array}$$

เมื่อ θ_1 แทน Trait Level เบื้องต้น

$\theta_2,....,\theta_M$ แทน Trait Level ที่ได้จากการปรับปรุงใหม่จากการวัดก่อนหน้านี้

แผนภาพที่ 2.2 การออกแบบกระบวนการทางโครงสร้างของ Wiener สำหรับ 4 เงื่อนไขของการวัด ตามโครงสร้างที่กำหนดไว้

จากแผนภาพที่ 2.2 คอลัมน์แสดงถึง Traits ส่วนแถวแสดงถึงเงื่อนไขของการวัดภายใต้การใช้ข้อสอบซึ่งเป็นสิ่งที่สามารถสังเกตได้ โดยค่าที่เท่ากับ 0 บ่งบอกถึงมิตินั้นไม่เกี่ยวข้องกับทักษะการปฏิบัติงาน แต่เมื่อพิจารณาถึง $\underline{\theta}_1$ ซึ่งเป็น Trait Level ตั้งต้น พบว่าจะมีความเกี่ยวข้องกับทุกเงื่อนไขที่มุ่งวัด ส่วน $\underline{\theta}_2$ แทนการเปลี่ยนแปลงจากเงื่อนไขที่ 1 ไปยังสถานการณ์ที่ 2 เช่นเดียวกับ $\underline{\theta}_3$ และ $\underline{\theta}_4$ แทนการเปลี่ยนแปลงจากเงื่อนไข 2 ไปยังเงื่อนไขที่ 3 และจากเงื่อนไขที่ 3 ไปยังเงื่อนไขที่ 4 ตามลำดับ ซึ่งสามารถสรุปได้ว่ากระบวนการทางโครงสร้างของ Wiener เหมาะสำหรับข้อมูลที่มีค่าความแปรปรวนเพิ่มขึ้นข้ามเงื่อนไขห ซึ่งเกี่ยวข้องกับลักษณะที่ต้องการวัดมากขึ้น เช่น การวัดความรู้ความสามารถทางสมองที่มีความซับซ้อน จะทำให้ค่าความแปรปรวนเพิ่มขึ้นตลอดเวลา Embretson (1991)

การออกแบบโครงสร้างที่ต้องการวัดจะยึดโมเดลราสช์ ดังสมการ

$$P\left(x_{i(k)s}=1|\underline{\theta}_s,\beta_i\right)=\frac{\exp\left(\sum_m \theta_{sm}+\beta_i\right)}{1+\exp\left(\sum_m^k \theta_{sm}-\beta_i\right)} \quad \text{.....(2.14)}$$

เมื่อ θ_{sm} แทน ระดับความสามารถของผู้สอบที่ j ของลักษณะ (Trait) ที่ m

β_i แทน ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i

จากสมการที่ (2.14) เห็นได้ว่า สัญลักษณ์ $\sum_m^k$ เป็นการควบคุมเพื่อให้เกิดความเหมาะสมสำหรับ $\theta_2,....,\theta_M$ ซึ่งเป็น Trait Level ที่ดัดแปลงขึ้นใหม่ภายใต้เงื่อนไขที่ k ซึ่งการนำโมเดล MRCML มาใช้ในการประมาณค่าสำหรับโมเดลที่มีลักษณะพหุมิติ จะเกิดความคลาดเคลื่อนมาตรฐานทั้งความสามารถตั้งต้นและ Trait Level ที่ได้ดัดแปลงขึ้นใหม่

ภายใต้เงื่อนไขที่มีความชัดเจน มีความเป็นไปได้ว่า การแผ่ขยายแนวคิดของโมเดล UIRT สำหรับใช้ในการวัดการเปลี่ยนแปลงในมิติที่มุ่งวัด เช่น Fischer (1997) ได้ขยายโมเดล Linear Logistic

Latent Trait Model (LLTM) ไปใช้ในกรณีแบบพหุมิติ โดยการระบุเทคนิคของผู้สอบที่ (Technical Persons) โดยผู้สอบกลายเป็นเทคนิคอย่างหนึ่งที่มีความแตกต่างกันตั้งแต่ 2 คุณลักษณะขึ้นไป ในแต่ละเงื่อนไข อย่างไรก็ตาม LLTM ไม่ใช่โมเดลของ MIRT และค่าความคลาดเคลื่อนมาตรฐานในการเปลี่ยนแปลงของ Trait Level สำหรับแต่ละบุคคลไม่ได้มีการเปลี่ยนแปลงเป็นปกติ

(2.3) โมเดลในการระบุโครงสร้างของระดับคุณลักษณะ (Models With Specified Trait Level Structures)

เมื่อแบบสอบที่นำมาใช้อยู่ภายใต้เงื่อนไขที่มีความหลากหลาย จะมีความเกี่ยวข้องกันอย่างเป็นระบบกับ Trait Level โดยการออกแบบโครงสร้างมีความเกี่ยวข้องภายในกับโครงสร้าง Trait Level ตามโมเดล IRT ในการระบุลักษณะเงื่อนไขที่มีลักษณะเฉพาะ ซึ่งจากมุมมองต่างๆ ของการออกแบบโครงสร้างปกติ มีการกำหนด Trait Level ออกเป็นส่วนย่อยๆ เพื่อนำผลมาเปรียบเทียบกับเงื่อนไขโดยรวมเพื่อวัดทักษะการปฏิบัติ โดยปัจจุบันพบว่า โมเดล IRT หลายๆ โมเดลมีจุดมุ่งหมายของการออกแบบโครงสร้างที่ต่างกันไป เช่น Embretson (1995, 1997) ได้นำเสนอ Structured Latent Traits Model (SLTM) โดยมีจุดมุ่งหมายในการออกแบบโครงสร้างของความสามารถด้านต่างๆ ไม่เฉพาะแต่ค่าอำนาจจำแนกของข้อสอบ แต่รวมถึงความสัมพันธ์ภายในของข้อสอบภายใต้เงื่อนไขการวัดที่แตกต่างกัน

ในปีเดียวกัน Wang, Wilson & Adams ได้เสนอโมเดลเชิงโครงสร้างในแต่ละมิติที่คล้ายคลึงกับโมเดลราสช์ ซึ่งเป็นโมเดลที่สามารถอ้างอิงไปยังสถานการณ์โดยทั่วไปและสามารถนำไปใช้ร่วมกับการออกแบบโครงสร้างอื่นๆ เช่น ร่วมกับโมเดล SLTM นอกจากนี้ Wang, Wilson & Adams (1995) ได้เสนอโปรแกรมทางคอมพิวเตอร์ MATS มาใช้ในการประมาณค่าพารามิเตอร์ ซึ่งคล้ายกับการศึกษาของ DiBello, Stout, & Roussos (1995) ที่ได้เสนอ Unified Model เป็นโมเดลที่มีการรวมความสามารถและการออกแบบโครงสร้างของข้อสอบเข้าด้วยกัน ถือได้ว่าเป็นแนวคิดที่นิยมใช้โดยทั่วไป นอกจากนี้พบว่าไม่มีผลกระทบสำหรับข้อจำกัดของการประยุกต์ใช้โมเดลราสช์ที่กำหนดให้ค่าอำนาจจำแนกของข้อสอบเป็นค่าคงที่

การแสดงโมเดลการออกแบบโครงสร้าง จะพิจารณาจากโมเดล SLTM ซึ่งเป็นส่วนประกอบหนึ่งของสมการ Multidimensional Logistic Model ดังสมการ

$$P\left(x_{i(k)s}=1\middle|\theta_s,\beta_i,\lambda_{i(k)m}\right)=\frac{\exp\left(\sum_m \lambda_{i(k)m}\theta_{sm}+\sum_k \beta_{ik}\right)}{1+\exp\left(\sum_m^k \lambda_{i(k)m}\theta_{sm}-\sum_k \beta_{ik}\right)} \quad \text{.....(2.15)}$$

เมื่อ θ_{sm} แทน ระดับความสามารถของผู้สอบที่ s ของลักษณะ (Trait) ที่ m

β_{ik} แทน การแจกแจงค่าพารามิเตอร์ความยากของเงื่อนไขที่ k สำหรับข้อสอบที่ i

$\lambda_{i(k)m}$ แทน ค่าน้ำหนักที่มีความเฉพาะเจาะจงของลักษณะ (Trait) ที่ m ของข้อสอบข้อที่ i ภายใต้เงื่อนไขที่ k

ความสำคัญของโมเดล SLTM จะแตกต่างจาก Multidimensional Logistic Model ในสมการที่ (2.15) โดยการให้น้ำหนัก จะมุ่งไปที่การกำหนดค่าเป็น 0 ดังนั้น โมเดล SLTM จึงคล้ายกับ Confirmatory Factor Analysis Model มากกว่า Exploratory Factor Analysis Model เช่น การออกแบบ MRMLC เป็นกรณีเฉพาะของ โมเดล SLTM แม้ว่าโมเดลดังกล่าวจะคล้ายคลึงกับรูปแบบโมเดล 2 พารามิเตอร์ ซึ่งจะแตกต่างจาก Exploratory Multidimensional IRT Model โดย SLTM กำหนดให้ค่าอำนาจจำแนกข้อสอบเป็นค่าคงที่ สามารถแสดงเมตริกซ์ ของการออกแบบMRMLC ซึ่งแทนด้วย Λ_{Kxm} ดังแสดงในแผนภาพ

$$\Lambda_{Kxm} = \begin{matrix} \lambda_{11} & 0 & 0 & 0 \\ \lambda_{21} & \lambda_{21} & 0 & 0 \\ \lambda_{21} & \lambda_{21} & \lambda_{21} & 0 \\ \lambda_{21} & \lambda_{21} & \lambda_{21} & \lambda_{21} \end{matrix}$$

เมื่อ $\underline{\lambda}_{km}$ แทนค่าอำนาจจำแนกของข้อสอบของเงื่อนไขที่ k ภายใต้มิติที่ m

แผนภาพที่ 2.3 เมตริกซ์ ของการออกแบบ MRMLC ซึ่งเป็นรูปแบบหนึ่งของโมเดล SLTM

ค่าคงที่ของค่าอำนาจจำแนกข้อสอบจะมี 2 ชนิด สำหรับการออกแบบ MRMLC โดยรูปแบบแรกเป็นค่าที่กำหนดให้เท่ากับ 0 เมื่อเป็นค่าอำนาจจำแนกในมิติเดียวกัน ส่วนรูปแบบที่สองจะมีค่าไม่เท่ากับ 0 เมื่อศึกษาข้ามมิติ

(2.4) โมเดลสำหรับการจำแนกกลุ่มบุคคล (Models for Distinct Classes of Persons)

โมเดล IRT เป็นโมเดลที่ศึกษากลุ่มคุณลักษณะแฝงที่ไม่สามารถสังเกตได้โดยตรงจากการศึกษาตัวแปรภายนอก โดยการแบ่งกลุ่มของลักษณะต่างๆ จะเป็นกระบวนการอย่างเป็นระบบที่แตกต่างกันของรูปแบบการตอบข้อสอบ ซึ่งโมเดลจะประกอบไปด้วย Class และ Trait Parameter เพื่อใช้ในการทำนายลักษณะของการตอบข้อสอบ แม้ว่าเทคนิควิธีของโมเดล IRT จะเป็นการศึกษาเพียง 1 Trait Level ของผู้สอบแต่ละคน แต่สามารถนำมาใช้ในการจำแนกเชิงพหุมิติได้ เช่น โมเดล SALTUS ที่พัฒนาโดย Wilson (1985) มีความเหมาะสมกับการจำแนกความแตกต่างของความรอบรู้กับไม่รอบรู้ การเพิ่มขึ้นของความน่าจะเป็นในการตอบข้อสอบได้ถูกมักจะได้รับอิทธิพลมาจากขั้นของการวัดที่มีความเฉพาะเจาะจง โดยสามารถแสดงได้ดังสมการ

$$P\left(x_{is}=1|\underline{\theta}_s,\beta_i\right)=\prod_m \frac{\exp\left(\theta_s-\beta_i+\xi_{h(s)k(i)}\right)}{1+\exp\left(\theta_s-\beta_i+\xi_{h(s)k(i)}\right)} \quad \text{.....(2.16)}$$

เมื่อ $\xi_{h(s)k(i)}$ แทน การเพิ่มขึ้นของโอกาสในการตอบข้อสอบได้ถูกสำหรับข้อสอบในชนิดที่ k ในขั้นที่ h

สิ่งสำคัญประการหนึ่งของการเลือกใช้โมเดล SALTUS คือ ICC จะขึ้นอยู่กับระดับขั้นของการพัฒนา ซึ่งต่อ Rost (1990) ได้นำเสนอโมเดล Mixed Population Rasch Model (MIRA) โดยเหมาะกับข้อมูลในการแบ่งกลุ่มความแตกต่างของบุคคลอย่างเป็นระบบ ซึ่งเป็นพื้นฐานของการตอบข้อสอบ การมีโครงสร้างหรือลำดับขั้นการเรียนรู้ที่ต่างกัน จะสามารถจำแนกรูปแบบของค่าความยากของข้อสอบในแต่ละกลุ่ม ถือได้ว่าเป็นโมเดลที่ใช้ในการประมาณค่า Trait Level และการจำแนกกลุ่มผู้สอบในแต่ละบุคคล

ในปี ค.ศ. 1997 Von Davier มีการพัฒนาโปแกรมทางด้านคอมพิวเตอร์ที่เรียกว่า WINMIRA เพื่อประมาณค่าความยากของข้อสอบภายในกลุ่มที่ทำการจัดจำแนก ซึ่งในปัจจุบันพบว่าได้มีการพัฒนาอย่างต่อเนื่องเกี่ยวกับโมเดล ในการจำแนกผู้สอบออกเป็นกลุ่มความสามารถและนำมาประยุกต์ใช้กันอย่างกว้างขวางในปัจจุบัน

กล่าวโดยสรุป โมเดล MIRT สามารถแบ่งได้เป็น 2 ประเภท คือ Exploratory MIRT Model และ Confirmatory MIRT Model ซึ่งมีลักษณะคล้ายคลึงกับการวิเคราะห์องค์ประกอบ โดย Exploratory MIRT Model เกี่ยวข้องกับการประมาณค่าพารามิเตอร์ข้อสอบและผู้สอบที่มากกว่า 1 มิติ เพื่อให้โมเดลมีความสอดคล้องกับข้อมูล โดยไม่ได้กำหนดจำนวนองค์ประกอบล่วงหน้า ประกอบด้วย 2 ประเด็นย่อย คือ 1) Multidimensional Logistic Model และ 2) Multidimensional Normal Ogive Model โดยโมเดลย่อยทั้ง 2 มีสูตรในการประมาณค่าคล้ายคลึงกับ Logistic Model และ Normal Ogive Model ในโมเดล UIRT แต่จะต่างกันในจำนวนมิติที่ศึกษา ซึ่งตรงข้ามกับ Confirmatory MIRT Model เกี่ยวข้องกับการประมาณค่าพารามิเตอร์ในมิติที่มีความเฉพาะเจาะจง โดยคล้ายกับการวิเคราะห์องค์ประกอบ ที่เกี่ยวข้องกับการระบุความสัมพันธ์ของข้อสอบไปยังมิติต่างๆ โดยพบว่า ข้อสอบจะประกอบไปด้วยการออกแบบตามโครงสร้างและเป็นโมเดลในรูปฟังก์ชันเชิงคณิตศาสตร์ที่มีการเชื่อมโยงกับคุณลักษณะที่มีความเฉพาะเจาะจงในหลายลักษณะ โมเดลนี้เหมาะสมสำหรับข้อสอบหรือชิ้นงานของทักษะการปฏิบัติ ที่ได้มีกำหนดหรือสร้างทฤษฎีไว้ล่วงหน้า ซึ่ง Embretson & Reise (2000) แบ่งออกเป็นโมเดลย่อยทั้งหมด 4 โมเดล คือ 1) Models for Noncompensatory Dimensional, 2) Models for Learning and Change, 3) Models With Specified Trait Level Structures และ 4) Models for Distinct Classes of Persons โดย Confirmatory MIRT Model เหมาะสำหรับการนำไปประยุกต์ใช้ได้หลากหลาย เช่น แก้ปัญหาทางด้านการวัดทางจิตวิทยาในการวัดความเปลี่ยนแปลง (Measurement of Change) ออกแบบโครงสร้างของความสามารถด้านต่างๆ เป็นต้น

นอกจากการศึกษาประเภทของโมเดล MIRT แล้ว สิ่งสำคัญประการหนึ่งคือความถูกต้องในการระบุมิติที่ทำการศึกษาและการประมาณค่าพารามิเตอร์ ในรายละเอียดต่อไป ได้นำเสนอประเด็นเกี่ยวกับการตรวจสอบมิติแฝง และการประมาณค่าพารามิเตอร์ตามโมเดล MIRT โดยมีรายละเอียดดังนี้

1.4 การตรวจสอบมิติแฝง (Assessing Dimensional)

ก่อนที่จะกล่าวถึงการตรวจสอบมิติแฝง ผู้วิจัยขอนำเสนอมโนทัศน์เบื้องต้นของมิติของแบบสอบ เพื่อให้เข้าใจความหมายของคำว่า มิติ และลักษณะของมิติเพื่อเป็นแนวคิดสำคัญสู่ขั้นตอนที่ถูกต้องของการตรวจสอบมิติของแบบสอบในขั้นตอนต่อไป

1.4.1 มโนทัศน์เกี่ยวกับมิติของแบบสอบ (Concept of Dimensional)

คุณภาพอย่างหนึ่งของเครื่องมือ คือการวัดได้ตรงตามสิ่งที่ต้องการวัด แต่เดิมมาในการวัดผลทางการศึกษา วิธีการสอบได้รับความนิยมอย่างแพร่หลาย คือ การสอบข้อเขียน การสอบข้อเขียนมักใช้แบบความเรียง (Essay Type) หรือแบบอัตนัย ซึ่งก็พบกันว่าการทำข้อสอบแบบความเรียงนั้น ผู้ตอบต้องใช้ความรู้ความสามารถหลายๆ ด้านพร้อมกัน จนทำให้ยากที่จะทำให้ระบุได้ว่า ข้อสอบนั้นวัดความสามารถเฉพาะด้านใดด้านหนึ่ง ประกอบกับการตรวจข้อสอบประเภทนี้ต้องใช้เวลาค่อนข้างมาก และอาจความเป็นปรนัยทำให้เกิดปัญหาคุณภาพเครื่องมือ นอกจากนี้ยังมีข้อจำกัดในด้านความครอบคลุมเนื้อหาที่ต้องการวัด (สุวิมล ติรกานันท์, 2548)

เมื่อเป็นเช่นนี้ นักวัดผลจึงได้คิดรูปแบบข้อข้อสอบปรนัยขึ้นเพื่อแก้ปัญหาดังกล่าว ได้แก่ ข้อสอบแบบตอบสั้น แบบเติมคำ แบบถูกผิด แบบจับคู่ และแบบเลือกตอบ โดยเฉพาะแบบเลือกตอบ เป็นแบบเลือกตอบเป็นแบบที่ได้รับความนิยมในขณะนี้ ในส่วนของการให้คะแนนนั้น แบบสอบปรนัยสามารถให้คะแนนแบบ 0-1 (Dichotomous-Scored Item) คือข้อที่ตอบผิดจะได้ 0 คะแนน และข้อที่ตอบถูกจะได้ 1 คะแนน ลักษณะของข้อสอบแบบปรนัยจะใช้เวลาในการตรวจน้อยลง สามารถใช้เครื่องจักรกลช่วยในการตรวจได้ และทำให้รายงานผลได้รวดเร็ว แม้จะมีผู้สอบจำนวนมากก็ตาม อีกทั้งยังสามารถสร้างข้อสอบให้ครอบคลุมเนื้อหาที่กำหนดได้ รวมทั้งสะดวกต่อการพัฒนาเป็นแบบสอบมาตรฐาน ดังนั้นความสำคัญของจำนวนมิติที่ศึกษาเป็นสิ่งที่นักวัดผลควรให้ความสำคัญ ซึ่งความหมายของคำว่า “มิติของแบบสอบ” ในมุมมองทางด้านการวัดผลจะมุ่งประเด็นไปยังโครงสร้างของแบบสอบที่ต้องการวัด เมื่อผู้วิจัยได้ทำการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องพบว่าความหมายของความเป็นมิติไม่มีผู้กล่าวไว้ชัดเจน แต่จะกล่าวความหมายของความเป็นเอกมิติ และความเป็นพหุมิติของแบบสอบ อย่างไรก็ตามก่อนที่จะกล่าวถึงความหมายของความเป็นเอกมิติและพหุมิติของแบบสอบ ควรที่จะทราบความหมายของคำว่า มิติ (Dimensional) ว่ามีความหมายอย่างไร เมื่อผู้จัยทำการสังเคราะห์ สามารถสรุปความหมายของมิติของแบบสอบ ได้ดังนี้

มิติของแบบสอบ (Dimension of Test) หมายถึง ลักษณะของความเป็นตัวแทนทางทฤษฎีภายใต้คุณลักษณะแฝง (Trait) แนวความคิด คุณลักษณะ (Attribute) กระบวนการ (Process) โครงสร้าง (Structure) โดยใช้แบบสอบซึ่งเป็นชุดของข้อสอบเป็นเครื่องมือวัดดังกล่าว (Messick, 1989) ซึ่งโดยทั่วไปสามารถแบ่งลักษณะมิติของแบบสอบได้เป็น 2 ประเภทที่แตกต่างกัน คือ

(1) แบบสอบที่มีลักษณะการวัดคุณลักษณะเด่นเพียงคุณลักษณะเดียว เรียกว่า “ความเป็นเอกมิติของแบบสอบ”

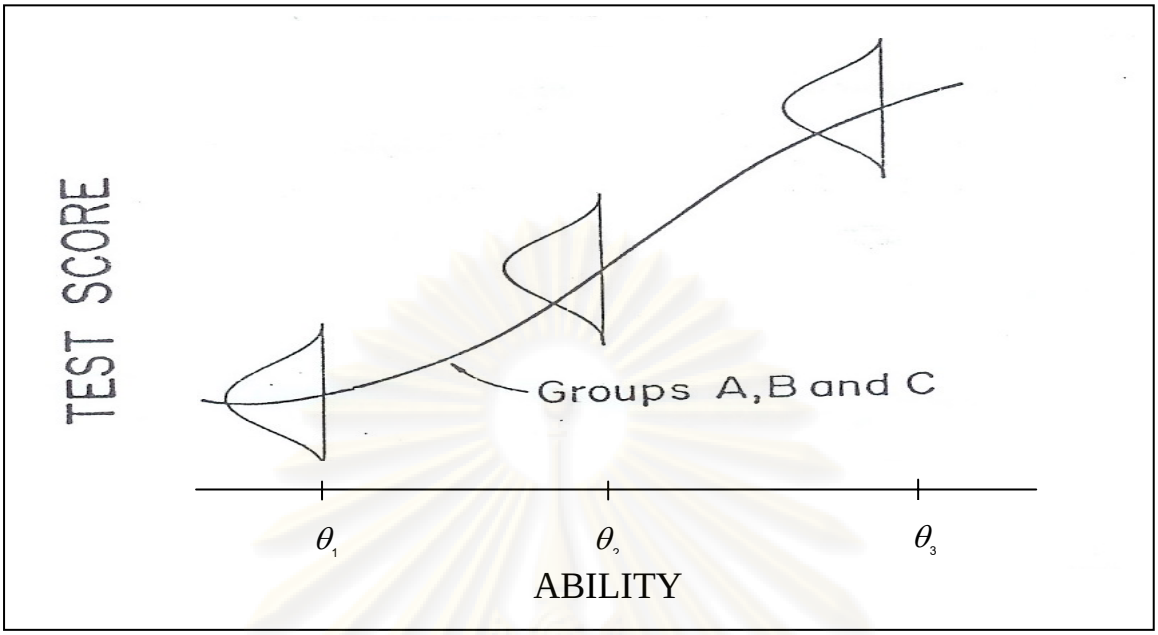
(2) แบบสอบที่มีลักษณะการวัดคุณลักษณะเด่นได้หลายคุณลักษณะในคราวเดียวกัน เรียกว่า “ความเป็นพหุมิติของแบบสอบ”

จากลักษณะของมิติของแบบสอบดังกล่าวข้างต้นสามารถให้ความหมายในแต่ละลักษณะ ได้ดังต่อไปนี้

ความเป็นเอกมิติของแบบสอบ (Unidimensional) ซึ่งมีความหมายในเชิงทฤษฎีตามแนวคิดของ McDonald (1999) หมายถึง ลักษณะของโครงสร้าง (Construct) หรือคุณลักษณะภายในมีลักษณะเป็นองค์ประกอบเชิงเดี่ยว (Factorial Simple) ที่วัดเพียงมิติเดียว เช่น แบบสอบทางคณิตศาสตร์ที่วัดโครงสร้างด้านพีชคณิต (Algebra) เพียงมิติเดียว เป็นต้น

สำหรับความหมายของ **ความเป็นเอกมิติของแบบสอบ** ในเชิงปฏิบัติตามแนวคิดของการวิจัยในครั้งนี้จะยึดนิยามตามมโนทัศน์ของคุณลักษณะแฝง (Latent Trait Theory) โดยใช้การวิเคราะห์โครงสร้างคุณลักษณะแฝง (Latent Structure Analysis) ซึ่งในโมเดลของคุณลักษณะแฝงเราจะสมมติ (assume) ว่าผู้เข้าสอบได้ตอบสนองต่อข้อสอบที่มีการตรวจให้คะแนนเป็น 0-1 และพิสูจน์แล้วว่าข้อสอบเหล่านี้เป็นอิสระจากกันทางสถิติ (Mutually Statistic Analysis Independent) ถ้าเพียง 1 คุณลักษณะแฝง อธิบายการแจกแจง (Distribution) ของรูปแบบการตอบสนองของข้อสอบ n ข้อได้ แสดงว่าชุดของข้อสอบนี้มีความเป็นเอกมิติ (อ้างใน วรนุช แหยมแสง, 2536)

ความเป็นเอกมิติของข้อมูลที่สมบูรณ์แบบหาไม่ได้ แม้ว่าจะมีแบบสอบที่วัดความสามารถเพียงมิติเดียวก็ตาม แต่เนื่องจากผู้ตอบข้อสอบไม่ได้ใช้ความสามารถเพียงด้านเดียวในการตอบสนองต่อแบบสอบฉบับนั้น ทั้งนี้มักจะมีองค์ประกอบอื่นเข้ามาเกี่ยวข้อง เช่น พุทธพิสัยอื่นๆ บุคลิกภาพ และองค์ประกอบในการสอบ (Test Taking Factor) ที่มีผลต่อการทำแบบสอบ องค์ประกอบเหล่านี้รวมถึงระดับแรงจูงใจ ความกังวล ความสามารถในการทำงานเร็ว ความรู้เกี่ยวกับกระดาษคำตอบ ทักษะทางพุทธิพิสัยต่างๆ และรวมถึงคุณลักษณะเด่น (Dominant) รวมอยู่ในการวัดครั้งเดียวด้วยชุดของข้อสอบชุดนั้นด้วย ดังนั้นสิ่งที่ข้อตกลงเบื้องต้นนี้ต้องการก็เป็นเพียงแต่ให้ได้ข้อมูลจากชุดข้อสอบ ที่ให้ “องค์ประกอบเด่น 1 องค์ประกอบ” ที่เป็นผลมาจากการทำข้อสอบชุดนั้น และได้องค์ประกอบเด่นตัวนี้ คือ ความสามารถวัดได้ และเป็นลักษณะที่ผู้วัดต้องการวัด ส่วนองค์ประกอบอื่นๆ ที่เหลือซึ่งมีความแปรปรวนน้อยมาก เมื่อเทียบกับองค์ประกอบแรก จะเป็นส่วนเกิน ซึ่งสามารถแสดงการแจกแจงของคะแนนของแบบสอบที่มีลักษณะเป็นเอกมิติ โดยสมมติว่าแบบสอบฉบับหนึ่งประกอบด้วยข้อสอบ n ข้อ และนำไปสอบกับกลุ่มประชากรย่อยกลุ่มต่างๆ จำนวน r กลุ่ม แล้วพิจารณาการแจกแจงของคะแนนจากแบบสอบว่าแต่ละระดับความสามารถของกลุ่มผู้สอบเหล่านั้นมีการแจกแจงเหมือนกันหรือไม่ ถ้ามีการแจกแจงเหมือนกันแสดงว่าแบบสอบนั้นมีความเป็นเอกมิติ ดังแผนภาพที่ 2.4



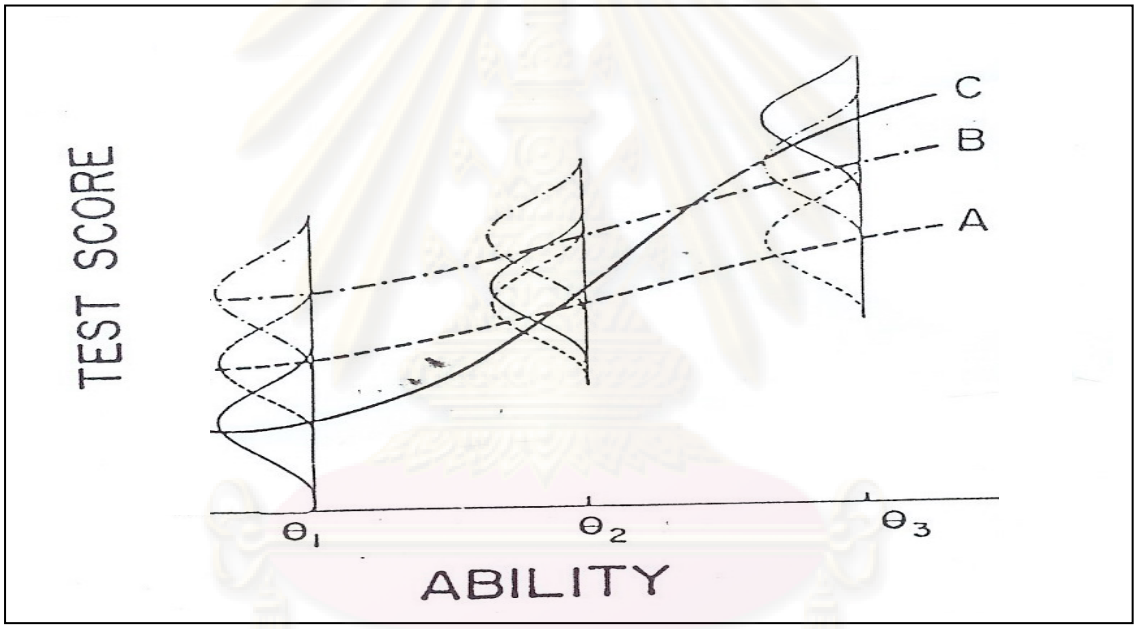
แผนภาพที่ 2.4 การแจกแจงคะแนนของแบบสอบที่มีลักษณะเป็นเอกมิติ (Unidimensional) (อ้างใน วรนุช แหยมแสง, 2536)

อย่างไรก็ตามสำหรับสถานการณ์จริงของแบบสอบที่ใช้ ส่วนใหญ่ค่อนข้างเป็นการวัดที่ซับซ้อน และเนื้อหาการวัดในปัจจุบันเป็นการเรียนการสอนที่ต้องมีการบูรณาการข้ามเนื้อหาวิชาดังนั้นจึงมีความจำเป็นอย่างยิ่งที่ต้องศึกษาความหมายของความเป็นพหุมิติของแบบสอบ ซึ่งสามารถสรุปความหมายเชิงทฤษฎีและเชิงปฏิบัติ ได้ดังนี้

ความเป็นพหุมิติของแบบสอบ (Unidimensional) ซึ่งมีความหมายในเชิงทฤษฎีตามแนวคิดของ McDonald (1999) หมายถึง ลักษณะของโครงสร้าง (Construct) หรือคุณลักษณะภายในมีลักษณะเป็นองค์ประกอบเชิงซ้อน (Factorials Complex) ภายใต้การวัดหลายมิติ (McDonald, 1999) ตัวอย่างเช่น แบบสอบคณิตศาสตร์ชุดหนึ่งอาจประกอบไปด้วยการวัดพีชคณิต (Algebra) และเรขาคณิต (Geometry) ซึ่งในกรณีนี้ ชุดของแบบสอบย่อยมิติแรกกำหนดให้เป็นการวัดทางด้านพีชคณิต ส่วนมิติที่ 2 เป็นการวัดทางด้านเรขาคณิต ซึ่งเป็นการศึกษาในลักษณะ 2 มิติ เป็นต้น

สำหรับความหมายของ **ความเป็นพหุมิติของแบบสอบ** ในเชิงปฏิบัติการตามแนวคิดของการวิจัยในครั้งนี้จะยึดนิยามตามมโนทัศน์ของคุณลักษณะแฝง (Latent Trait Theory) โดยใช้การวิเคราะห์โครงสร้างคุณลักษณะแฝง (Latent Structure Analysis) ซึ่งในโมเดลของคุณลักษณะแฝงสมมติให้ (assume) ผู้เข้าสอบได้ตอบสนองต่อข้อสอบที่มีการตรวจให้คะแนนเป็น 0-1 และพิสูจน์แล้วว่าข้อสอบเหล่านี้เป็นอิสระจากกันทางสถิติ (Mutually Statistic Analysis Independent) ถ้ามีมากกว่า 1 คุณลักษณะแฝงที่สามารถอธิบายการแจกแจง (Distribution) ของรูปแบบการตอบสนองของข้อสอบ n ข้อ แสดงว่าชุดของข้อสอบนี้มีความเป็นพหุมิติ (อ้างใน วรนุช แหยมแสง, 2536) โดย Zhang & Stout (1999) ได้ระบุว่า รูปแบบที่แน่นอนของการแบ่งแยกกลุ่มของข้อสอบนั้นเมื่อนำแบบสอบย่อย

มารวมกัน (Test Composite) ควรมีความเป็นตัวแทนกับการกำหนดแบบแผนแบบสอบ (Test Specification) โดยแนะนำให้วิเคราะห์มิติหรือการวิเคราะห์ข้อสอบตามโมเดล MIRT เพื่อตรวจสอบหรือสนับสนุนสิ่งที่ต้องการวัดได้ ในการตัดสินใจในส่วนประกอบต่างๆ โดยเฉพาะในบริบทของการวิเคราะห์เนื้อหาหรือทักษะความสามารถทางด้านสมอง ที่พิจารณาจากการตอบข้อสอบได้ถูกต้อง ซึ่งจะเป็นประโยชน์อย่างยิ่งในการวิเคราะห์ชุดของแบบสอบย่อย เพื่อใช้จำแนกมิติที่จะวัด ดังนั้น ขั้นตอนแรกของการวิเคราะห์ MIRT คือ การกำหนดข้อมูลเมื่อข้อมูลมีลักษณะหลายมิติ ซึ่งสามารถแสดงการแจกแจงของคะแนนของแบบสอบที่มีลักษณะเป็นพหุมิติ โดยสมมติว่าแบบสอบฉบับหนึ่งประกอบด้วยข้อสอบ n ข้อ และนำไปสอบกับกลุ่มประชากรย่อยกลุ่มต่างๆ จำนวน r กลุ่ม แล้วพิจารณาการแจกแจงของคะแนนจากแบบสอบว่าแต่ละระดับความสามารถของกลุ่มผู้สอบเหล่านั้น มีการแจกแจงเหมือนกันหรือไม่ ถ้าการแจกแจงมีการแปรเปลี่ยนในแต่ละระดับความสามารถแสดงว่าแบบสอบนั้นวัดความสามารถมากกว่า 1 ความสามารถ ดังแสดงในแผนภาพดังแผนภาพที่ 2.5



แผนภาพที่ 2.5 การแจกแจงคะแนนของแบบสอบที่มีลักษณะเป็นพหุมิติ (Multidimensional) (อ้างใน วรนุช แหยมแสง, 2536)

1.4.2 วิธีการตรวจสอบมิติของแบบสอบ (Assessing Dimensional)

แต่เดิมนั้น การวิเคราะห์องค์ประกอบแบบเส้นตรง จะใช้ความสัมพันธ์แบบเตตระคลอริค (Tetrachoric Correlation) และใช้ความสัมพันธ์แบบหลายค่า (Polytomous) สำหรับกรณีที่ข้อมูลมีลักษณะเป็นแบบหลายค่า (Polytomous) เพื่อใช้ในการประเมินมิติการวัดของแบบสอบ แต่มักเกิดปัญหาในกรณีที่ความสัมพันธ์ระหว่างข้อสอบภายใต้คุณลักษณะแฝงมีลักษณะไม่ใช่เส้นตรง (Hattie, 1984) ซึ่งมีสาเหตุจากการความไม่เหมาะสมกันระหว่างโมเดลและข้อมูลเชิงประจักษ์ มิติต่างๆ อาจมีการปะปนกับค่าความยากของข้อสอบ เช่น องค์ประกอบที่เป็นตัวแทนของข้อสอบหลายๆ ข้อ สามารถเปรียบเทียบระดับความยากของข้อสอบ และสามารถจำแนก

ออกเป็นความแตกต่างของมิติได้ นอกจากนี้การวัดทางด้านคณิตศาสตร์ใช้วิธีการวิเคราะห์องค์ประกอบจำเป็นต้องใช้เมตริกซ์ความสัมพันธ์แบบเตตระคลอริค ซึ่งมีลักษณะความสัมพันธ์ทางบวก และเงื่อนไขนี้อาจไม่ชัดเจน เมื่อนำมาใช้กับข้อมูลจริง อาจกล่าวได้ว่าไม่มีวิธีการใดที่เป็นมาตรฐานในการกำหนดจำนวนขององค์ประกอบได้อย่างมีความหมาย (Mislevey, 1986 อ้างใน Ackerman, Gierl & Walker, 2003)

วิธีการศึกษาเชิงประจักษ์หลายๆ วิธี ที่มีจุดมุ่งหมายในการตรวจสอบมิติแบบสอบ เช่น จากการศึกษาของ Hatti (1984; 1985) การศึกษาของ Hambletion & Rovinelli (1986) รวมถึงการศึกษา Theory of Essential Unidimensional ของ Stout (1987) ซึ่งศึกษาภายใต้การสอบแบบ Nonparametric ตั้งอยู่บนข้อตกลงเบื้องต้นของการตรวจสอบคุณสมบัติเด่นเพียงคุณลักษณะเดียวที่มีอิทธิพลต่อการตอบในแต่ละข้อของผู้สอบ (Hatti, Krakowski, Roger, & Swaminathan, 1996; Nandakumar, 1991; Stout et al., 1996; Zhang & Stout, 1999) อย่างไรก็ตาม การวิเคราะห์โดยส่วนใหญ่จะเป็นเชิงสำรวจ ซึ่งผลที่ได้จากการตรวจสอบมิติอาจจะขัดแย้งกับสมมติฐานที่ตั้งไว้ล่วงหน้า

แนวทางสำคัญในการตัดสินผลการตรวจสอบมิติของแบบสอบ สามารถสรุปได้เป็น 3 ขั้นตอน (Ackerman, Gierl, & Walker, 2003) คือ

1) กำหนดแบบแผนของแบบสอบ (Test Specification) เพื่อเป็นแนวทางในการเขียนข้อสอบตามเนื้อหาและหน่วยที่ต้องการวัดที่แสดงถึงความเป็นตัวแทนของมิติที่มุ่งวัดได้อย่างถูกต้อง

2) การวิเคราะห์เนื้อหา (Content Analysis) เพื่อเป็นแนวทางในการตรวจสอบมิติ ซึ่งมี 2 วิธี วิธีแรกอาจให้ผู้เชี่ยวชาญที่มีประสบการณ์ในการตรวจสอบหรือตัดสินมิติที่ต้องการวัด หรือวิธีที่สองการตัดสินโดยการยึดเนื้อหาเป็นหลัก

3) การวิเคราะห์ทางจิตวิทยา (Psychological Analysis) เพื่อเป็นแนวทางในการตรวจสอบมิติในกรณีที่โครงสร้างแบบสอบได้กำหนดตามสมมติฐานหรือเป็นแนวคิดหรือทัศนะทางด้านจิตวิทยา เช่น การวิเคราะห์ชิ้นงานเพื่อวัดความสามารถทางสมอง สามารถบ่งบอกถึงคุณลักษณะของทักษะทางด้านคณิตศาสตร์ที่ต้องการวัด (Gallagher, 1998; Gallagher, De Lisi, Holst, McGillcuddy De Lisi, Morely, & Cahalan, 2000) ซึ่งความสามารถทางด้านสมองเหล่านี้สามารถบ่งบอกได้ โดยการดำเนินการใช้ชุดข้อสอบเพื่อใช้ในการตรวจสอบมิติได้

โดยทั่วไปการตรวจสอบมิติแฝงจะมีเทคนิคเฉพาะในการตรวจสอบ ตลอดจนมีการพัฒนาโปรแกรมคอมพิวเตอร์มาใช้ในการตรวจสอบ การเลือกใช้วิธีการตรวจสอบมิติ สิ่งหนึ่งที่ต้องตระหนัก คือ กระบวนการที่ใช้เป็นเพียงเครื่องมือเท่านั้น แต่ไม่ใช่สิ่งที่ใช้ในการตัดสินทั้งหมด เนื่องจากการศึกษาที่ผ่านมาโดยส่วนใหญ่จะใช้ข้อมูลที่จำลองขึ้นมา มากกว่าการใช้ข้อมูลจริง ซึ่งการศึกษาโดยใช้ข้อมูลจริงยังมีจำนวนน้อย ในที่นี้ผู้วิจัยขอนำเสนอเทคนิคในการตรวจสอบมิติตามลำดับการพัฒนาที่ใช้กันโดยทั่วไปมี 4 วิธี คือ 1) TESTFACT 2) Hierarchical Cluster Analysis 3) DETECT และ 4) DIMTEST โดยมีรายละเอียดดังต่อไปนี้

1) TESTFACT

สุวิมิล ติรการนันท์ (2548) ได้กล่าวถึงวิธีการพิจารณาความเป็นมิติของแบบสอบจากการวิเคราะห์องค์ประกอบเชิงยืนยัน จากโปรแกรม TESTFACT ว่า Wilson et al. (1991) ได้พัฒนาโปรแกรม TESTFACT ขึ้นเพื่อใช้ในการวิเคราะห์ข้อสอบและทดสอบความตรงของโครงสร้าง ด้วย χ^2 สำหรับ Likelihood Ratio (G^2) ในการตรวจสอบมิติของแบบสอบ ดัชนีที่ใช้นี้ทดสอบด้วยการกำหนดจำนวนองค์ประกอบของชุดข้อมูลไว้ล่วงหน้า แล้วทดสอบด้วย χ^2 ที่ประมาณค่าด้วยวิธี G^2 เพื่อทดสอบความเหมาะสมของโมเดล เมื่อค่า G^2 ไม่มีนัยสำคัญแสดงว่าข้อมูลมีจำนวนองค์ประกอบเท่าที่กำหนดในการทดสอบ ซึ่งคำนวณได้จากสมการ

$$G^2 = 2\sum_{l=1}^{2^n} rl \ln \frac{rl}{N\tilde{P}l} \quad \text{.....(2.17)}$$

เมื่อ rl แทน จำนวนความถี่ใน Pattern l

$\tilde{P}$ l แทน Item Parameter ได้จากการประมาณค่าด้วยวิธี Maximum Likelihood

i แทน จำนวนข้อมูล

ค่า degrees of freedom = $2^n(m+1)+\frac{m(m-1)}{2}$ .....(2.18)

เมื่อ m แทน จำนวนองค์ประกอบ

n แทน จำนวนข้อสอบ

2) Hierarchical Cluster Analysis

ในปี ค.ศ. 1992 Roussos ได้พัฒนาโปรแกรมคอมพิวเตอร์ 2 โปรแกรมคือ CCPROX และ HCA เพื่อใช้ในการแบ่งกลุ่มของตัวแปร ซึ่งกระบวนการนี้ใช้แบ่งกลุ่มจำนวนข้อสอบซึ่งมีลักษณะคล้ายคลึงกันให้อยู่ในกลุ่มเดียวกัน โปรแกรมสามารถวิเคราะห์ข้อสอบที่มีการตรวจให้คะแนนแบบ 2 ค่า ได้ถึง 120 ข้อ โดยไม่มีข้อจำกัดในเรื่องจำนวนผู้สอบ ซึ่งก่อนที่จะใช้เทคนิค Cluster Analysis จะต้องมีการวัดค่าความคล้ายกันของข้อสอบแต่ละคู่ที่เป็นไปได้ทั้งหมด โดยโปรแกรม CCPROX อนุญาตให้ผู้ใช้โปรแกรมสามารถเลือกใช้วิธีการวัดค่าความคล้ายกันของข้อสอบแต่ละคู่ แตกต่างกันได้ เช่นการประมาณค่าความแปรปรวนร่วมระหว่างข้อสอบแต่ละคู่ (Douglas, Kim, Roussos, Stout, & Zhang, 1999) สำหรับในสถานการณ์ของการจำลองข้อมูล จะมีการวัดความคล้ายคลึงกัน โดยการพิจารณาความไวต่อความเป็นพหุมิติ (Douglas et al., 1999; Hartz, Roussos, & Stout, 2000)

เมตริกซ์ความคล้ายคลึงกัน (Proximity Metrix) เป็นปัจจัยป้อน (Input) ที่สำคัญในโปรแกรม HCA เป็นผลจากความหลากหลายในการจัดกลุ่มข้อสอบ โดยกระบวนการเริ่มต้นในการใช้เทคนิค Cluster Analysis นั้น แต่ละข้อจะมีรูปแบบที่กำหนดไว้แล้ว ซึ่งมีขั้นตอนในการคำนวณทั้งหมด k-1 รอบ เมื่อกำหนดให้มีข้อสอบทั้งหมด k ข้อ โดยแต่ละรอบของการคำนวณ 2 กลุ่มใดหรือข้อสอบใดที่มีความสัมพันธ์กันหรือคล้ายกันมากที่สุดจะจัดอยู่ในกลุ่มเดียวกัน ส่วนขั้นตอนสุดท้ายข้อสอบทั้งหมดที่คล้ายกันจะถูกรวมให้มีเพียงกลุ่มเดียว (Single Cluster) ซึ่งโปรแกรม HCA

จะมีค่าสั่งในการกำหนดความคล้ายกันในการแบ่งกลุ่ม (Proximity Cluster) ผลที่ได้จากการจำลองข้อมูลมีข้อเสนอแนะว่าการใช้วิธีไม่ถ่วงน้ำหนักของค่าเฉลี่ยแต่ละคู่ในกลุ่ม (Unweighted Pair Group Method of Averages; UPGMA) เป็นวิธีการในการจำแนกหรือแบ่งกลุ่มชุดของข้อสอบที่มีความคล้ายคลึงกันของโครงสร้างที่ไม่ซับซ้อน (Approximate Simple Structure; APSS) (Douglas et al., 1999)

Hierarchical Cluster Analysis เป็นเทคนิคสำคัญสำหรับนักวิจัยที่มีข้อสงสัยในความเป็นพหุมิติของชุดแบบสอบ ซึ่งไม่แน่ใจว่าอยู่ภายใต้โครงสร้างใดโครงสร้างหนึ่งหรือไม่ อย่างไรก็ตามการวิเคราะห์ด้วยเทคนิคดังกล่าวผลที่ได้จะเป็นการจำแนกกลุ่มที่มีความแตกต่างกันโดยไม่คำนึงว่ามีลักษณะเป็นพหุมิติหรือไม่ ดังนั้นการพิจารณาตัดสินในแต่ชุดของแบบสอบย่อยจำเป็นต้องตีความจากผลลัพธ์ของโปรแกรม (Output) นอกจากนี้วิธีในการหมุนซ้ำ (Iteration) จนสำเร็จหรืออยู่ในระดับที่น่าพอใจนั้น ขึ้นอยู่กับวิธีการก่อนหน้านี้เป็นสำคัญ ซึ่งความเป็นไปได้ของผลลัพธ์ (Outcome) อาจจะช่วยอธิบายให้ข้อมูลมีความถูกต้องมากยิ่งขึ้น

3) DETECT

เป็นกระบวนการตรวจสอบมิติแฝงเชิงยืนยันแบบ Nonparametric ซึ่งจะใช้ในการประมาณค่าจำนวนของมิติแฝงที่มีคุณลักษณะเด่นในชุดของข้อมูลและสามารถตรวจสอบความเป็นเอกมิติของแบบสอบ โดยทั่วไปโปรแกรม DETECT จะระบุคุณลักษณะเด่นของมิติแฝงในแต่ละข้อ (Roussos, Reese, & Harris, 1997) โดยโปรแกรมสามารถวิเคราะห์ได้ถึง 120 ข้อคะแนนที่มีการตรวจให้คะแนนแบบ 0-1 กับผู้สอบไม่เกิน 6,000 คน โดยกระบวนการนี้เป็นการสร้างลักษณะที่เหมือนกันจัดอยู่ในกลุ่มเดียวกัน โดยผู้ใช้โปรแกรมสามารถระบุจำนวนมิติแฝงสูงสุดที่ต้องการศึกษาได้ เนื่องจากการจัดกลุ่มชุดของข้อสอบ แต่กระบวนการดังกล่าวยังมีลักษณะแบบไม่เป็นทางการเท่าใดนัก เนื่องจากการระบุการจัดกลุ่มเพื่อจำแนกความแตกต่างของมิติจะอาศัยกระบวนการในการระบุความเป็นหนึ่งเดียว (Zhang & Stout, 1999)

4) DIMTEST

วิธีการตรวจสอบมิติแฝงที่ได้รับความนิยมมากที่สุดวิธีหนึ่งในปัจจุบันคือ การวิเคราะห์ด้วยโปรแกรม DIMTEST (Froelich, 2000; Froelich & Habing, 2001; Stout, Froelich, & Gao, 2001) ในเวอร์ชันใหม่ล่าสุด ซึ่งเป็นกระบวนการ Nonparametric Statistical เพื่อตรวจสอบสมมติฐานของแบบสอบ คล้ายคลึงกับการตรวจสอบด้วยโปรแกรม DIMTEST ก่อนหน้านี้ที่จะทำการตรวจสอบความสัมพันธ์ระหว่างชุดข้อสอบย่อยภายใต้เงื่อนไขความแปรปรวนร่วมของข้อสอบ แต่จะแตกต่างจากการศึกษาก่อนหน้านี้โดยนำไปสู่การวิเคราะห์เชิงยืนยัน โดยโปรแกรมสามารถวิเคราะห์ได้ถึง 120 ข้อคะแนนที่มีการตรวจให้คะแนนแบบ 0-1 กับผู้สอบไม่เกิน 6,000 คน การวิเคราะห์ไม่ยุ่งยากมากนัก เพียงแต่ผู้ใช้เลือกชุดของข้อสอบย่อยในแบบสอบ เพื่อวัดมิติแฝงที่มีคุณลักษณะเด่นเหมือนกัน และเลือกวิธีที่จะใช้ในการตรวจสอบ เช่น Substantive Judgment, Hierarchical Cluster Analysis หรือ DETECT เป็นต้น

สถิติที่ใช้ทดสอบ คือ ค่า T ซึ่งสามารถคำนวณได้จากโปรแกรม DIMTEST เพื่อแสดงถึงระดับความแตกต่างของมิติของ 2 ชุดข้อสอบย่อย ซึ่งค่า T จะถูกกำหนดให้มีการแจกแจงเป็นปกติกับ

ค่าที่คาดหวัง (Expected Value) เท่ากับ 0 ภายใต้สมมติฐานที่ตั้งไว้ก่อนหน้านี้ เช่น มีการกำหนดสมมติฐานไว้ล่วงหน้าว่าประกอบด้วยการวัดความสามารถทางคณิตศาสตร์ 2 มิติ คือ มิติที่ 1 วัดด้านมิติสัมพันธ์ (Spatial) และ มิติที่ 2 วัดด้านทั่วไป (General) โดยกำหนดให้ข้อที่ 1-6 เป็นชุดของแบบสอบในมิติที่ 1 มีความแตกต่างจากข้อที่เหลือในแบบสอบ ที่เป็นชุดของข้อสอบในมิติที่ 2 ผลที่ได้จากการวิเคราะห์ค่า T = 2.69 (p=.004) แสดงว่าชุดของข้อสอบในมิติที่ 1 แตกต่างจากชุดของข้อสอบในมิติที่ 2 นั่นคือแบบสอบทางด้านคณิตศาสตร์ฉบับนี้มีการศึกษาใน 2 มิติ เป็นต้น

มีผู้แนะนำไว้ว่า ผลที่ได้จากการใช้โปแกรม DIMTEST ของชุดข้อสอบที่วัดด้านมิติสัมพันธ์ (Spatial) จะถูกระบุโดยใช้การศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องเพื่อนำมาใช้ในการศึกษาความแตกต่าง เมื่อมีการเปรียบเทียบกับชุดของข้อสอบที่เหลืออีกชุดหนึ่ง อย่างไรก็ตาม ผลที่ได้จากการวิเคราะห์จะบ่งชี้ว่า ชุดของข้อสอบที่วัดด้านมิติสัมพันธ์ (Spatial) ค่อนข้างให้ค่าที่น้อย ซึ่งบางทีการรวมข้อสอบข้อที่ 1-6 อาจจะไม่พบการจัดกลุ่มกัน เมื่อนำมาวิเคราะห์ด้วย Hierarchical Cluster Analysis

กล่าวโดยสรุป วิธีการที่ได้นำเสนอมาทั้ง 4 วิธี จะให้มุมมองที่แตกต่างกันในการตรวจสอบมิติแฝง โดยทั่วไปการศึกษาทางด้านพหุมิตินั้น ผลที่ได้จากการใช้กระบวนการดังกล่าวจะขึ้นอยู่กับความหลากหลายของความสามารถที่ต้องการวัดซึ่งเป็นลักษณะเชิงโครงสร้างที่นักวิจัยได้กำหนดขึ้นมา แต่สิ่งสำคัญอย่างหนึ่งคือ วิธีการทั้ง 4 วิธีจะถูกออกแบบเพื่อตรวจสอบโครงสร้างมิติความสามารถ ที่ยังไม่สามารถระบุได้ว่าวิธีการใดที่ดีที่สุดซึ่งในประเด็นนี้จำเป็นที่จะต้องมีการศึกษาและพัฒนาต่อไปในอนาคต

จากที่กล่าวมาทั้งหมดเกี่ยวกับโมเดล MIRT ซึ่งเป็นแนวคิดพื้นฐานในการนำไปประยุกต์ใช้ให้เกิดผลในทางปฏิบัติต่อไป โดยเป็นประเด็นสำคัญมากที่สุดประการหนึ่งในการนำโมเดล MIRT ไปประยุกต์ในการเชื่อมโยงคะแนนระหว่างแบบสอบเพื่อแก้ไขจุดอ่อนที่เคยศึกษาในอดีตดังที่ได้กล่าวมาแล้วข้างต้น (Rackase, 1997) ในรายละเอียดต่อไปผู้วิจัยได้นำเสนอแนวคิดเกี่ยวกับการเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติรวมถึงงานวิจัยที่เกี่ยวข้อง เพื่อเป็นพื้นฐานสำคัญสำหรับการศึกษาในกระบวนการเชื่อมโยงคะแนนตามแบบพหุมิติตามโมเดล MIRT ในการนำไปสู่ระเบียบวิธีวิจัยของผู้วิจัย โดยมีรายละเอียดดังนี้

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตอนที่ 2 แนวคิดเกี่ยวกับการเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Item Response Theory Linking; MIRT Linking) และงานวิจัยที่เกี่ยวข้อง

2.1 มโนทัศน์เบื้องต้นของการเชื่อมโยงคะแนน (Principle Concept of Linking Scores)

การศึกษาถึงวิธีการเชื่อมโยงคะแนนแบบพหุมิติ พบว่า มีการนำเอาแนวคิดของโมเดล MIRT มาเป็นองค์ประกอบสำคัญในการศึกษา ซึ่งก่อนที่จะมีการกล่าวถึงแนวคิดของการเชื่อมโยงคะแนนตามโมเดล MIRT จำเป็นอย่างยิ่งที่ต้องเข้าใจถึงมโนทัศน์เบื้องต้นของการเชื่อมโยงคะแนน เนื่องจากศัพท์คำว่าการเชื่อมโยงคะแนนเป็นคำที่มีความหมายหมายกว้าง ครอบคลุมรูปแบบของการเชื่อมโยงคะแนนทั้งหมด ซึ่งถ้าหากไม่สามารถเข้าใจถึงความแตกต่างของรูปแบบการเชื่อมโยงคะแนนในแต่ละวิธี อาจทำให้เกิดมโนทัศน์ที่คลาดเคลื่อน อันนำไปสู่การนำรูปแบบการเชื่อมโยงคะแนนไปใช้ในสถานการณ์ที่ไม่เหมาะสม และผิดไปจากหลักความเป็นจริง ดังนั้นในหัวข้อนี้ผู้วิจัยจึงนำเสนอถึง กรอบแนวคิดของการเชื่อมโยงคะแนนระหว่างแบบสอบสองฉบับ เงื่อนไขที่ใช้เป็นเกณฑ์ในการจำแนกรูปแบบการเชื่อมโยงคะแนน รายละเอียดในการจำแนกรูปแบบในการเชื่อมโยงคะแนน และปัจจัยที่ส่งผลต่อความตรงของการเชื่อมโยงคะแนนจากแบบสอบ 2 ฉบับ เพื่อเป็นพื้นฐานสำคัญในการศึกษาถึงวิธีการเชื่อมโยงคะแนนแบบพหุมิติ โดยมีรายละเอียดดังนี้

2.1.1 กรอบแนวคิดของการเชื่อมโยงคะแนนระหว่างแบบสอบสองฉบับ

แต่เดิมนักวัดผลทางการศึกษาสามารถนำคะแนนมาปรับเทียบกันได้ด้วยวิธีการทางสถิติที่เรียกว่า “การปรับเทียบคะแนนระหว่างแบบสอบ (Test Score Equating)” ซึ่งถือเป็นกระบวนการออกแบบการทดสอบและเทคนิคการวิเคราะห์ผลการสอบ เพื่อให้สามารถนำผลการสอบจากแบบสอบต่างฉบับและผู้สอบต่างกลุ่มกัน มาเปรียบเทียบกันอย่างเที่ยงธรรม โดยมีข้อตกลงเบื้องต้นเกี่ยวกับแบบสอบทั้งสองฉบับ คือ วัดคุณลักษณะเด่นเดียวเหมือนกัน (Unidimensionality) มีลักษณะของความเป็นคู่ขนานในด้านเนื้อเรื่อง โครงสร้าง รูปแบบ ชนิดข้อสอบ เวลาที่ใช้สอบ ความเที่ยงสูงทัดเทียมกัน (ศิริชัย กาญจนวาสี, 2550)

Flanagan (1951, อ้างใน Kolen, 2004) เป็นผู้อภิปรายอย่างลึกซึ้งเกี่ยวกับการเชื่อมโยงของคะแนนและการสร้างมาตรวัดคะแนนโดยใช้คำว่า ความสามารถในการเปรียบเทียบ (Comparability) ที่ใช้ในการเชื่อมโยงคะแนนจากแบบสอบที่มีการกระจายของคะแนนสอบที่มีความคล้ายคลึงกันจากประชากรผู้สอบเดียวกัน โดย Flanagan ทำการจำแนกกระบวนการของการเชื่อมโยงคะแนนจากแบบสอบสองฉบับที่วัดคุณลักษณะเดียวกันและแบบสอบสองฉบับที่วัดคุณลักษณะต่างกัน ต่อมา Angoff (1971, อ้างใน Kolen, 2004) ทำการสรุปและจัดทำการเชื่อมโยงคะแนนและการวัดให้เป็นระบบมากขึ้น โดยใช้คำว่า “การปรับเทียบ (Equating)” เพื่อใช้ในการอธิบายการเชื่อมโยงคะแนนระหว่างแบบสอบหลายฉบับ โดยมีการสอบคนละครั้งและแบบสอบถูกสร้างจากการกำหนดคุณลักษณะเฉพาะของข้อสอบเดียวกัน และมีความคิดเห็นที่ไม่สอดคล้องกับ Flanagan ว่าความเป็นจริงแล้ว ผลของคะแนนที่ได้จากการการปรับเทียบควรมีคุณสมบัติของความเป็นอิสระจากกลุ่มผู้สอบ นอกจากนี้ Angoff ได้นิยามศัพท์คำว่า“การทำคะแนนให้เป็นมาตรฐาน (Calibration)” ว่าเป็น

การเชื่อมโยงคะแนนจากแบบสอบที่วัดคุณลักษณะเดียวกันแต่มีความยาก/ความเที่ยงต่างกัน เช่น การเชื่อมโยงคะแนนของแบบสอบวัดผลสัมฤทธิ์ระดับประถมศึกษาที่ถูกออกแบบมาเพื่อใช้ในระดับชั้นต่าง ๆ ซึ่งสามารถทำให้เป็นมาตรฐานได้โดยเรียกว่า การสร้างสเกลแนวตั้ง (Vertical Scaling) นอกจากนี้ Angoff กล่าวว่า *"การเชื่อมโยงคะแนนที่ได้จากแบบสอบที่วัดคุณลักษณะต่างกัน ผลของคะแนนที่ได้จะผันแปรตามกลุ่มผู้สอบ"*

สำหรับรูปแบบต่างๆ ที่นำมาใช้ในการเชื่อมโยงระหว่างแบบสอบสองฉบับ มีผู้เชี่ยวชาญหลายท่านได้ทำการจำแนก โดยพิจารณาตามหลักเกณฑ์ที่แตกต่างกัน ดังที่จะกล่าวถึงในรายละเอียดต่อไปเกี่ยวกับเงื่อนไขที่ใช้เป็นเกณฑ์ในการจำแนกรูปแบบการเชื่อมโยงคะแนน

2.1.2 เงื่อนไขที่ใช้เป็นเกณฑ์ในการจำแนกรูปแบบการเชื่อมโยงคะแนน

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการเชื่อมโยงคะแนนในรูปแบบต่างๆ และชนิดของคะแนนที่ได้จากการเชื่อมโยงระหว่างแบบสอบสองฉบับ สามารถสรุปถึงแนวคิดในการแบ่งชนิดของการเชื่อมโยงคะแนนและผลของคะแนนที่ได้จากการเชื่อมโยงคะแนนดังนี้

Mislevey (1992) ได้เสนอแนวคิดเกี่ยวกับการวัดในกรณีที่แบบสอบสองฉบับมีการวัดที่ไม่มีคุณลักษณะของความร่วมกันหรือไม่เกี่ยวข้องกัน และ Linn (1993) ได้นำแนวคิดดังกล่าวมาใช้ในการจำแนกประเภทของการเชื่อมโยงคะแนนด้วยวิธีต่างๆ โดยนำวิธีการทางสถิติเป็นเครื่องมือในการจัดกลุ่มและจำแนกรูปแบบการเชื่อมโยงคะแนน โดยพิจารณาจากความแกร่งของผลคะแนนที่ได้จากการเชื่อมโยง และมีเกณฑ์เกี่ยวกับการจำแนก คือ พิจารณาว่าแบบสอบมีเนื้อหาที่ใช้วัดที่มีลักษณะเดียวกันหรือไม่และค่าสถิติของแบบสอบทั้งสองฉบับมีลักษณะเดียวกันหรือไม่ ซึ่งสามารถจำแนกได้เป็น 4 ชนิด คือ การปรับเทียบคะแนน (Equating) การทำให้เป็นคะแนนมาตรฐาน (Calibration) การฉายภาพ (Projection)และการปรับด้วยค่าทางสถิติ (Statistical Moderation)

นอกจากนี้ Feuer et al. (1999) ได้เสนอเกณฑ์การจำแนกรูปแบบการเชื่อมโยงคะแนนเพิ่มเติม โดยจำแนกการเชื่อมโยงคะแนนระหว่างแบบสอบตามแบบแผนในการสร้างและพัฒนาแบบสอบ โดยยังคงไว้ซึ่งรูปแบบในการเชื่อมโยงทั้ง 4 ชนิดดังกล่าวข้างต้น ที่เป็นไปตามแนวคิดของ Mislevy (1992) & Linn (1993) ซึ่งเกณฑ์ที่ Feuer et al. (1999) ใช้ในการจำแนกแบ่งเป็น 3 ขั้นตอน คือ

ขั้นตอนที่ 1 การกำหนดโครงร่างของแบบสอบ (Framework Definition) สามารถประเมินจากขั้นตอนการกำหนดโครงร่างของแบบสอบทั้งสองฉบับ โดยพิจาณาจากการระบุถึงขอบเขต (Scope) ปริมาณ (Extent) ของกรอบมวลพฤติกรรม (Domain) เพื่อทำให้ทราบว่าผู้เรียนมีความสามารถอะไรบ้างและมีความรู้ระดับใด

ขั้นตอนที่ 2 การกำหนดคุณลักษณะเฉพาะของแบบสอบ หรือการสร้างผังข้อสอบ (Test Specification or Blueprint) เป็นการเสนอรายละเอียดของการทดสอบแต่ละครั้งว่าวัดเนื้อหาอะไร หรือจุดมุ่งหมายเชิงพฤติกรรมที่ต้องการวัดจากผู้เรียน ขอบเขตของเนื้อหาวิชา ตลอดจนมีการกำหนดน้ำหนักความสำคัญ/สัดส่วนของจำนวนข้อกระทง/ข้อสอบที่จะสร้างขึ้นเป็นแบบสอบที่ใช้

สำหรับวัดพฤติกรรมตามขอบเขตของเนื้อหาวิชาที่ต้องการทดสอบแต่ละครั้ง และรวมถึงการตรวจให้คะแนนด้วย

ขั้นตอนที่ 3 การคัดเลือกข้อสอบ (Item Selection) เป็นการพิจารณาจากข้อสอบที่ถูกคัดเลือกมาบรรจุไว้ในแบบสอบ ข้อสอบที่คัดเลือกมาควรเป็นตัวแทนที่ดีของผังข้อสอบให้มากที่สุดเท่าที่จะเป็นไปได้

โดยปกติในการกำหนดโครงร่างของแบบสอบ (Framework Definition) ส่วนมากจะมีการกำหนดในลักษณะที่เป็นส่วนหนึ่งของกรอบมวลพฤติกรรม (Domain) เช่น ถ้าต้องการวัดกรอบมวลพฤติกรรมด้านการอ่านของผู้เรียน ฉะนั้นในการสร้างแบบสอบควรมีการกำหนดโครงร่างให้ได้ตามกรอบมวลพฤติกรรมที่กำหนดไว้ให้ได้มากที่สุด

จากที่กล่าวมาข้างต้นเกี่ยวกับกรอบแนวคิดของกรอบมวลพฤติกรรมของแบบสอบสามารถสรุปได้ตามแนวคิดของ Feuer et al. (1999) ที่ได้เสนอเกณฑ์จำแนกการเชื่อมโยง 3 ลักษณะ คือ

เกณฑ์ที่ 1 แบบสอบทั้งสองฉบับมีการกำหนดโครงร่างของแบบสอบและมีการกำหนดคุณลักษณะเฉพาะของแบบสอบเดียวกัน

เกณฑ์ที่ 2 แบบสอบทั้งสองฉบับมีการกำหนดโครงร่างของแบบสอบที่มีลักษณะเดียวกันแต่มีการกำหนดคุณลักษณะเฉพาะของแบบสอบที่ต่างกัน

เกณฑ์ที่ 3 แบบสอบทั้งสองฉบับมีการกำหนดโครงร่างของแบบสอบและมีการกำหนดคุณลักษณะเฉพาะของแบบสอบที่ต่างกัน

กรอบแนวคิดเกี่ยวกับเกณฑ์ในการจำแนกการเชื่อมโยงคะแนนของ Feuer et al. (1999) สามารถสรุปได้ดังตาราง

ตารางที่ 2.2 การจำแนกประเภทของการเชื่อมโยงคะแนนตามเกณฑ์ของ Feuer et al. (1999)

ประเภทของการเชื่อมโยงคะแนน	การกำหนดโครงร่างของแบบสอบ (Framework)	การกำหนดคุณลักษณะเฉพาะของแบบสอบของแบบสอบ (Test Specifications)
การปรับเทียบคะแนน (Equating)	มีลักษณะเดียวกัน	มีลักษณะเดียวกัน
คะแนนความสอดคล้อง (Concordance/Moderation)	มีลักษณะเดียวกัน	มีลักษณะแตกต่างกัน
การใช้แบบสอบเป็นตัวปรับอยู่ภายใน (Moderation)	มีลักษณะแตกต่างกัน	มีลักษณะแตกต่างกัน

จากที่กล่าวมาข้างต้นผู้วิจัยจะนำเสนอรายละเอียดเกี่ยวกับรูปแบบการเชื่อมโยงคะแนนของ Mislevy (1992) & Linn (1993) ผนวกกับเกณฑ์ในการจำแนกของ Feuer et al. (1999) ซึ่งสามารถจำแนกรูปแบบการเชื่อมโยงคะแนนได้เป็น 4 รูปแบบ คือ

รูปแบบที่ 1 การปรับเทียบคะแนน (Equating) เป็นการเชื่อมโยงคะแนนจากแบบสอบที่สอบคนละครั้ง แต่แบบสอบถูกสร้างมาด้วยลักษณะของความร่วมกันหรือมีความสัมพันธ์กันในด้านลักษณะของเนื้อหาและค่าสถิติของแบบสอบ (Angoff, 1971; Kolen, 2004) และมีคุณสมบัติของความไม่ผันแปรตามกลุ่มผู้สอบ

รูปแบบที่ 2 การทำให้เป็นคะแนนมาตรฐาน (Calibration) เป็นการเชื่อมโยงคะแนนจากแบบสอบที่มีลักษณะโครงสร้างเดียวกัน แต่แบบสอบทั้ง 2 ฉบับมีค่าความยากหรือความเที่ยงต่างกัน มีความแกร่งน้อยกว่าการปรับเทียบและไม่มีคุณสมบัติของความไม่ผันแปรตามกลุ่มผู้สอบ

รูปแบบที่ 3 การฉายภาพ (Projection) เป็นการเชื่อมโยงคะแนนจากแบบสอบที่วัดคุณลักษณะต่างกัน โดยใช้วิธีการทำนายเข้ามาช่วยในการเชื่อมโยงคะแนน

รูปแบบที่ 4 การปรับค่าทางสถิติ (Statistical Moderation) เป็นการเชื่อมโยงคะแนนจากแบบสอบที่วัดคุณลักษณะต่างกัน มีข้อตกลงเบื้องต้นเกี่ยวกับแบบสอบน้อยที่สุดและมีตัวแปรปรับ (Moderator) แฝงอยู่ในแบบสอบแต่ละฉบับ เช่น การวัดผลสัมฤทธิ์ทางภาษาไทยกับภาษาอังกฤษโดยมีแบบสอบ SAT I ที่วัดความสามารถด้านภาษาเป็นตัวแปรปรับระหว่างแบบสอบสองฉบับ เป็นต้น

2.1.3 รายละเอียดในการจำแนกรูปแบบในการเชื่อมโยงคะแนน

Mislevy (1992) ได้เสนอแนวคิดเกี่ยวกับการวัดในกรณีที่แบบสอบ 2 ฉบับ วัดคุณลักษณะต่างกัน และ Linn (1993) ได้นำแนวคิดดังกล่าวมาใช้ในการจำแนกชนิดของการเชื่อมโยงคะแนนระหว่างแบบสอบ 2 ฉบับที่วัดคุณลักษณะต่างกัน โดยเรียกว่า “รูปแบบในการเชื่อมโยงคะแนน” ซึ่งใช้วิธีการทางสถิติและความแกร่งจากผลของคะแนนที่ได้จากการเชื่อมโยงเป็นเครื่องมือในการจำแนก แบ่งออกเป็น 4 ชนิด คือ 1) การปรับเทียบคะแนน (Equating) 2) การทำคะแนนให้เป็นมาตรฐาน (Calibration) 3) การฉายภาพ (Projection) และ 4) การปรับด้วยค่าสถิติ (Statistical Moderation)

Mislevy (1992) & Linn (1993) กล่าวว่าผลของคะแนนที่ได้จากการเชื่อมโยงชนิดการปรับด้วยค่าสถิติ (Statistical Moderation) และการฉายภาพ (Projection/Prediction) มีความเป็นอิสระจากกลุ่มผู้สอบและการแปลงคะแนนด้วยวิธีการปรับด้วยค่าสถิติ (Statistical Moderation) มีคุณสมบัติของความสมมาตร (Symmetric) แต่วิธีการฉายภาพ (Projection/Prediction) ไม่มีคุณสมบัติของความสมมาตร (Symmetric)

รวมทั้ง Kolen & Brennan (2004) ได้จำแนกรูปแบบการเชื่อมโยงคะแนนตามเกณฑ์ของ Feuer et al. (1999) ผนวกกับแนวคิดของ Mislevy (1992) & Linn (1993) ในการจำแนกรูปแบบการเชื่อมโยงคะแนน ซึ่งมีรายละเอียดในการจำแนกรูปแบบการเชื่อมโยงคะแนนและเกณฑ์ที่ใช้ดังต่อไปนี้

รูปแบบที่ 1 การปรับเทียบคะแนน (Equating) เป็นการเชื่อมโยงคะแนนจากแบบสอบที่มีการสอบกันคนละครั้ง แต่แบบสอบถูกสร้างมาด้วยลักษณะของความร่วมกันหรือสัมพันธ์กันในด้านเนื้อหาและค่าสถิติของแบบสอบ หรือการมีลักษณะของเนื้อหาเดียวกันและมีค่าสถิติของแบบสอบแบบเดียวกัน

ถ้าพิจารณาในกรณีของการจำแนกรูปแบบตามเกณฑ์ของ Feuer et al. โดยใช้เกณฑ์ในการจำแนก คือ การกำหนดโครงร่างของแบบสอบและการกำหนดคุณลักษณะเฉพาะของแบบสอบ พบว่าแบบสอบทั้ง 2 ฉบับต้องมีการกำหนดโครงร่างของแบบสอบและการกำหนดคุณลักษณะเฉพาะของแบบสอบที่มีลักษณะเดียวกัน และควรมีคุณสมบัติของความไม่ผันแปรตามกลุ่ม คือ ไม่ว่าจะนำผลของคะแนนที่ได้ไปใช้จะไม่ขึ้นอยู่กับกลุ่มผู้สอบ

รูปแบบที่ 2 การทำให้เป็นคะแนนมาตรฐาน (Calibration) เป็นการเชื่อมโยงคะแนนจากแบบสอบที่วัดคุณลักษณะเดียวกันแต่แบบสอบทั้ง 2 ฉบับมีค่าความยากหรือค่าความเที่ยงต่างกัน ถ้าพิจารณาในกรณีของการจำแนกรูปแบบตามเกณฑ์ของ Feuer et al. โดยใช้เกณฑ์ในการจำแนก คือ การกำหนดโครงร่างของแบบสอบและการกำหนดคุณลักษณะเฉพาะของแบบสอบ พบว่าสามารถแบ่งได้เป็น 2 กรณี คือ 1) กรณีที่แบบสอบทั้ง 2 ฉบับที่จะทำเป็นคะแนนมาตรฐานได้นั้น จะต้องมีการกำหนดโครงร่างของแบบสอบที่มีลักษณะเดียวกันแต่มีการกำหนดคุณลักษณะเฉพาะของแบบสอบที่ต่างกัน และ 2) กรณีที่แบบสอบทั้ง 2 ฉบับมีการกำหนดโครงร่างของแบบสอบและการกำหนดแบบแผนแบบสอบต่างกัน ตลอดจนในกรณีที่อาจมีการใช้โครงร่างของแบบสอบร่วมกัน ดังตัวอย่างต่อไปนี้

ตัวอย่างที่ 1 การทำให้คะแนนเป็นมาตรฐานในกรณีที่แบบสอบ 2 ฉบับมีความร่วมกันในด้านเนื้อหาแต่มีค่าสถิติของแบบสอบที่ต่างกัน ซึ่งอาจเกิดเนื่องมาจาก ความยาวของแบบสอบ 2 ฉบับไม่เท่ากัน แบบสอบที่มีความยาวมากกว่าก็จะมีค่าความเที่ยงมากกว่าแบบสอบที่มีความยาวที่สั้นกว่า

ตัวอย่างที่ 2 การทำให้คะแนนเป็นมาตรฐานในกรณีที่แบบสอบ 2 ฉบับมีเนื้อหาบางส่วนต่างกันและอาจ มี/ไม่มีค่าสถิติของแบบสอบที่ต่างกัน ซึ่งสามารถนำคะแนนมาจากระดับชั้นที่ต่างกันแล้วนำคะแนนมาสร้างให้อยู่ในสเกลร่วมกัน ระหว่างแบบสอบ 2 ฉบับ โดยในกรณี เรียกว่า **“การสร้างสเกลแนวตั้ง” (Vertical Scaling)**

ตัวอย่างที่ 3 การทำให้คะแนนเป็นมาตรฐานโดยใช้วิธีทางสถิติ เช่น ใช้โมเดล IRT ทำให้คะแนนเป็นมาตรฐาน โดยนำข้อคำถามมาสร้างเป็นสเกลร่วมกันระหว่างแบบสอบ 2 ฉบับ และประมาณค่าความสามารถของผู้สอบและสามารถนำคะแนนมาเปรียบเทียบกันได้ ซึ่งการทำให้คะแนนเป็นมาตรฐานด้วยวิธีนี้ควรอยู่ในกรอบของข้อตกลงเบื้องต้นเกี่ยวกับแบบสอบว่า แบบสอบ 2 ฉบับ ต้องมุ่งวัดคุณลักษณะเดียวกัน และควรมีเนื้อหาของแบบสอบที่คล้ายคลึงกันมากที่สุด ถ้าแบบสอบทั้ง 2 ฉบับมีเนื้อหาลักษณะเดียวกัน การทำให้เป็นคะแนนมาตรฐานในกรณีนี้ก็คือ การปรับเทียบนั่นเอง

รูปแบบที่ 3 การฉายภาพ (Projection) เป็นการเชื่อมโยงคะแนนจากแบบสอบที่วัดคุณลักษณะต่างกัน โดยใช้วิธีการทำนายเข้ามาช่วยในการเชื่อมโยงคะแนน และมีข้อแตกต่างจากการปรับเทียบ(equating)และการทำให้เป็นคะแนนมาตรฐาน (Calibration) คือ 1) เป็นการศึกษาความสัมพันธ์แบบทางเดียว คือ ขาดคุณสมบัติของความสมมาตร นั่นคือ สหสัมพันธ์ระหว่าง x กับ y มีค่าไม่เท่ากับ ความสัมพันธ์ระหว่าง y กับ x 2) เป็นวิธีที่เหมาะสมกับกรณีของการออกแบบใน

การเก็บรวบรวม ข้อมูลแบบผู้สอบกลุ่มเดียว และ 3) เป็นวิธีที่ไม่คำนึงถึงคุณลักษณะของแบบสอบว่าต้องมีลักษณะเดียวกัน

รูปแบบที่ 4 การปรับค่าทางสถิติ (Statistical Moderation) หรือในบางครั้งเรียกว่า “การปรับให้การกระจายของคะแนนเท่าเทียมกัน (Distribution Matching)” โดยใช้กรณีที่มีการออกแบบการเก็บรวบรวมข้อมูลแบบผู้สอบกลุ่มเดียว รวมถึงการออกแบบการเก็บรวบรวมข้อมูลแบบใช้ผู้สอบกลุ่มสุ่มและแบบการใช้ผู้สอบกลุ่มที่ไม่ทัดเทียมกันด้วย

นอกจากนี้ Dorans (2000) ได้แบ่งชนิดของการเชื่อมโยงคะแนนเป็น 3 ชนิด คือ

1. การปรับเทียบคะแนน (Equating) เป็นการเชื่อมโยงคะแนนโดยที่ผลคะแนนที่ได้สามารถนำมาทดแทนกันได้อย่างสมบูรณ์ (Fully Exchangeable) โดยแบบสอบทั้ง 2 ฉบับถูกสร้างมาโดยมีการกำหนดคุณลักษณะเฉพาะของข้อสอบที่มีลักษณะเดียวกัน

การปรับเทียบคะแนนระหว่างแบบสอบ (Test Equating) เป็นศัพท์เฉพาะที่นักจิตมิติ (Psychometrician) นำมาใช้ในกระบวนการวัดและประเมินผล ในช่วงแรกนักวัดผลโดยส่วนมากไม่ได้ให้ความสนใจเท่าที่ควร จะมีการศึกษากันเฉพาะกลุ่มนักจิตมิติเท่านั้น (Kolen & Brennan, 1995) ต่อมาในปี ค.ศ. 1980 ได้มีเอกสารเกี่ยวกับเรื่องนี้เพิ่มมากขึ้นทำให้นักวัดผลเริ่มเห็นความสำคัญและนำมาใช้ให้เป็นประโยชน์กับการทดสอบ 15 ปี ต่อมาได้มีการเพิ่มจำนวนและความหลากหลายของโปรแกรมการสอบที่มีการใช้แบบสอบหลายฉบับ ผู้รับผิดชอบในการทำแบบสอบมีความต้องการที่จะอ้างอิงบทบาทของการปรับเทียบคะแนนเพื่อวิจารณ์การทดสอบ ความเคลื่อนไหวในการตรวจสอบและติดตามผลการจัดการศึกษา และประเด็นความยุติธรรมเกี่ยวกับการทดสอบที่ปรากฏชัดเจนขึ้นทำให้ศาสตร์ด้านการปรับเทียบคะแนนได้รับการศึกษาและพัฒนาขึ้นมาตามลำดับ

วัตถุประสงค์ของการปรับเทียบคะแนนเป็นตัวกำหนดแนวทางในการดำเนินการปรับเทียบคะแนน เพื่อให้เกิดประโยชน์ตามความต้องการ สามารถจำแนกได้เป็น 2 กลุ่ม กลุ่มแรกมีวัตถุประสงค์เพื่อนำผลที่ได้จากการปรับเทียบคะแนนจากแบบสอบต่างฉบับไปเป็นข้อมูลในการตัดสินผลการเรียนหรือผลการศึกษาร่วมกัน กรณีนี้ใช้กระบวนการปรับเทียบในแนวนอน (Horizontal Equating) ส่วนอีกกลุ่มหนึ่งมีวัตถุประสงค์เพื่อนำผลการปรับเทียบไปใช้ในการพิจารณาพัฒนาการทางการเรียนหรือการเปลี่ยนแปลงทางการศึกษา ใช้การปรับเทียบคะแนนแนวตั้ง (Vertical Equating)

คุณสมบัติที่สำคัญในการปรับเทียบคะแนนระหว่างแบบสอบสองฉบับ ต้องมีคุณลักษณะพิเศษบางประการทั้งแบบสอบและวิธีการปรับเทียบคะแนน ดังนี้ (ศิริชัย กาญจนวาสี, 2550)

1) วัดคุณลักษณะเด่นเดียวเหมือนกัน (Unidimensionality)

2) มีลักษณะของความเป็นคู่ขนานในด้าน

2.1) เนื้อเรื่อง

2.2) โครงสร้าง

2.3) รูปแบบ

2.4) ชนิดข้อสอบ

2.5) เวลาที่ใช้สอบ

3) ความเที่ยงสูงทัดเทียมกัน

นอกจากนี้ วิธีการการปรับเทียบคะแนนระหว่างแบบสอบจะต้องเป็นวิธีการที่มีคุณสมบัติที่สำคัญ คือ

(1) ความเสมอภาค (Equity)

(2) ความไม่ผันแปรตามกลุ่ม (Invariance Across Group)

(3) ความสมมาตร (Symmetry)

2. การสร้างสเกล (Scaling) เป็นการเชื่อมโยงคะแนนโดยแปลงคะแนนจากแบบสอบหนึ่งมาอยู่บนมาตรวัดของแบบสอบอีกฉบับหนึ่ง อาศัยการจัดตำแหน่งของผู้สอบในแต่ละแบบสอบ โดยที่แบบสอบ 2 ฉบับ มีการวัดคุณลักษณะที่คล้ายคลึงกันแต่มีการกำหนดคุณลักษณะฉพาะของข้อสอบต่างกัน และผลของคะแนนที่ได้ไม่สามารถนำมาทดแทนกันได้

3. การทำนาย (Prediction) เป็นการเชื่อมโยงคะแนนที่อยู่ภายใต้ข้อตกลงเบื้องต้นของแบบสอบที่ผ่อนคลายข้อตกลงเบื้องต้นเกี่ยวกับแบบสอบมากที่สุด แต่ผลของคะแนนไม่มีคุณลักษณะของความสมมาตร (Symmetric)

ตามแนวคิดของ Dorans (2000) เกี่ยวกับชนิดของผลของคะแนนที่ได้จากการเชื่อมโยงคะแนนวิธีต่าง ๆ มีดังนี้

1. คะแนนที่ได้จากการปรับเทียบ (Equating Scores) เป็นผลของคะแนนที่ได้มาจากการเชื่อมโยงคะแนนชนิดการปรับเทียบคะแนน (Test Equating)

Hollan & Rubin (1982, อ้างใน Dorans, 2004) ได้กล่าวว่า เป้าหมายของการปรับเทียบคะแนนคือ สามารถนำคะแนนมาใช้แทนกันได้ (Interchangable) และคะแนนจะสามารถมาใช้แทนกันได้ ถ้าผลคะแนนที่ได้จากการวัดมาจากแบบสอบที่วัดคุณลักษณะเดียวกัน และมีลักษณะการวัดแบบเดียวกัน โดยแบบสอบที่นำมาปรับเทียบอาจเป็นแบบสอบที่มีการสร้างคนละครั้งแต่ใช้เครื่องมือในการวัดแบบเดียวกัน เช่น ในการวัดความยาวของวัตถุ/สิ่งของส่วนใหญ่ใช้ไม้บรรทัดเป็นเครื่องมือในการวัดความยาว สามารถวัดออกมาเป็นหน่วย ตั้งแต่ เซนติเมตร นิ้ว ฟุตและเมตร ซึ่งสามารแปลงหรือปรับเทียบหน่วยวัดจากนิ้วให้กลายเป็นเซนติเมตรและเซนติเมตรให้กลายเป็นนิ้วได้ นั่นคือความหมายของการปรับเทียบนั่นเอง ส่วนความยาวหมายถึงคุณลักษณะที่ใช้วัดนั่นเอง

แต่ในทางปฏิบัติจะพบว่าเครื่องมือที่ใช้ในการวัด เช่น ไม้บรรทัดยังคงมีข้อจำกัดในการวัดทำให้มีการใช้หน่วยในการปรับเทียบได้ไม่สมบูรณ์ เช่น การวัดด้วยไม้บรรทัดที่วัดหน่วยเป็นมิลลิเมตรจะพบว่ามีความแม่นยำกว่าการวัดด้วยหน่วยเป็นนิ้ว จากที่กล่าวมาข้างต้น จะเห็นว่ามีข้อจำกัดในการใช้หน่วยการวัดแทนกันทำให้ไม่สามารถแลกเปลี่ยนคะแนนกันได้อย่างสมบูรณ์ เนื่องจากการวัดยังผนวกความคลาดเคลื่อนในการวัดด้วย

2. คะแนนความสอดคล้อง (Concordant Scores) เป็นผลของคะแนนที่ได้จากการเชื่อมโยงคะแนนชนิดการสร้างสเกล (Scaling) Dorans (2004) กล่าวว่าการออกแบบการเก็บรวบรวมข้อมูลและเทคนิคทางสถิติที่ใช้ในการสร้างคะแนนความสอดคล้อง สามารถใช้ได้ในลักษณะเดียวกับ

การปรับเทียบคะแนน โดยคะแนนความสอดคล้องเป็นการเชื่อมโยงคะแนนระหว่างแบบสอบที่มีลักษณะมาตรวัดต่างกันหรือการเชื่อมโยงคะแนนระหว่าง 2 สเกลนั่นเอง

ความแตกต่างที่ชัดเจนของคะแนนที่ได้จาก 2 สเกล โดยนำมาปรับเทียบให้อยู่บนมาตรวัดที่มีลักษณะร่วมกัน (Scaling) กับการปรับเทียบคะแนน (Equating) คือ ถ้าสิ่งที่วัดมีค่าความเที่ยงเท่ากันและมีผลเช่นเดียวกันในทุกกลุ่มย่อยของประชากรผู้สอบและการใช้คะแนนทดแทนกันได้ (Interchangeable) เป็นคุณสมบัติที่ได้จากการเชื่อมโยงคะแนนด้วยวิธีการของการปรับเทียบคะแนน (Equating) ส่วนการสร้างสเกล (Scaling) หรือคะแนนความสอดคล้อง (Concordance) จะไม่มีคุณสมบัติของการใช้คะแนนทดแทนกันได้ (Interchangeable)

สำหรับลักษณะเด่นที่ใช้ในการจำแนกระหว่างการปรับเทียบและการสร้างสเกล คือ คุณสมบัติของความสมมาตร (Symmetric) เช่น เมื่อนำคะแนนมาปรับเทียบกันระหว่างแบบสอบ X กับแบบสอบ Y คะแนนจากแบบสอบ X ที่ได้ 150 คะแนนจะเท่ากับแบบสอบ Y ที่ได้ 25 คะแนน เป็นต้น แต่ถ้าเป็นคะแนนที่ได้มาจากคะแนนความสอดคล้อง สามารถกล่าวได้เพียงว่าระดับของคะแนนจากแบบสอบ X และแบบสอบ Y อยู่ในระดับเดียวกันเนื่องจากแบบสอบทั้ง 2 ฉบับที่นำมาสร้างคะแนนความสอดคล้องมีการวัดโครงสร้างต่างกัน

3. คะแนนที่คาดหวัง (Expected Scores) เป็นผลของคะแนนที่ได้มาจากการเชื่อมโยงคะแนนชนิดการฉายภาพ (Projection)

กล่าวโดยสรุป จากกรอบแนวคิดเกี่ยวกับวิธีการเชื่อมโยงคะแนนและคุณสมบัติของผลคะแนนที่ได้จากการเชื่อมโยงด้วยวิธีต่าง ๆ ที่กล่าวมาข้างต้น ผู้วิจัยสามารถสรุป วิธีการในการเชื่อมโยงคะแนนระหว่างแบบสอบ 2 ฉบับ ประกอบด้วย 4 วิธี คือ 1) การปรับเทียบคะแนน (Equating) เป็นการเชื่อมโยงคะแนนเมื่อรูปแบบของแบบสอบมีเนื้อหาเดียวกันและมีค่าสถิติของแบบสอบมีลักษณะเดียวกัน ทำให้คะแนนที่ได้จากการปรับเทียบสามารถนำคะแนนมาทดแทนกันได้ (Angoff, 1971) 2) การทำคะแนนให้เป็นมาตรฐาน (Calibration) เป็นการเชื่อมโยงคะแนนเมื่อแบบสอบวัดคุณลักษณะเดียวกันแต่มีความแตกต่างกันในค่าความยากและค่าความเที่ยงของแบบสอบ เช่น การปรับเทียบคะแนนจากแบบสอบชนิดสั้นกับยาว (Angoff, 1971) 3) การปรับค่าทางสถิติ (Statistical moderation) เป็นการเชื่อมโยงคะแนนเมื่อแบบสอบวัดคุณลักษณะเดียวกัน (Mislevy, 1992 & Linn, 1993) และ 4) ความสอดคล้องของคะแนน (Concordance) เป็นการเชื่อมโยงคะแนนเมื่อแบบสอบวัดคุณลักษณะที่คล้ายคลึงกัน แต่มีการกำหนดคุณลักษณะเฉพาะของข้อสอบต่างกัน และผลของคะแนนที่ได้ไม่สามารถนำมาทดแทนกันได้

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

2.1.4 ปัจจัยที่ส่งผลต่อความตรงของการเชื่อมโยงคะแนนจากแบบสอบ 2 ฉบับ

การนำผลของคะแนนที่ได้จากแบบสอบต่างฉบับ ที่มีการวัดคุณลักษณะต่างกัน หรือมีเนื้อหาของแบบสอบที่ต่างกันนั้น ย่อมมีผลต่อความตรงของคะแนนที่ได้จากการเชื่อมโยง โดยผู้วิจัยขอกล่าวถึงแนวคิดที่เกี่ยวข้องกับความตรง ดังนี้

Feuer et al. (1999) ได้สรุปถึงปัจจัยที่ส่งผลต่อความตรงของการเชื่อมโยงคะแนน ดังนี้

1. ความคล้ายคลึงกันของเนื้อหา (Content) ค่าความยาก (Difficult) และรูปแบบของข้อสอบ (Item Formats)
2. ความสามารถในการเปรียบเทียบความคลาดเคลื่อนของคะแนนที่ได้จากการวัด
3. การบริหารจัดการแบบสอบ
4. ผลที่เกิดขึ้นจากการใช้แบบสอบ เช่น ความแม่นยำและความคงที่ของการเชื่อมโยงคะแนนในกลุ่มย่อย (Subgroup) ต่าง ๆ เป็นต้น

จากแนวคิดของ Feuer et al. (1999) ได้กล่าวถึงปัจจัยที่ส่งผลต่อความตรงของการเชื่อมโยงคะแนน โดยพิจารณาจากความคล้ายคลึงกันของเนื้อหาแบบสอบ ซึ่งต่อมา Kolen & Brennan (2004) มีการขยายแนวคิดเกี่ยวกับความคล้ายคลึงกันของแบบสอบว่า สำหรับคำนิยามของความคล้ายคลึงได้เสนอหลักเกณฑ์การเลือกชนิดการเชื่อมโยงคะแนน โดยพิจารณาจากความคล้ายคลึงกันของแบบสอบ ซึ่งสามารถจำแนกเป็น 4 ประเภท คือ

1. การนำผลของคะแนนมาสรุปอ้างอิง (Inferences) เป็นการตอบคำถามว่า “ขนาด (extent) ของความคล้ายคลึงกันระหว่างแบบสอบ 2 ฉบับ อยู่ในระดับใด?” นั่นคือ พิจารณาถึงความมีเป้าหมายในการวัดร่วมกันระหว่างแบบสอบสองฉบับว่าอยู่ในระดับใด
2. โครงสร้าง/คุณลักษณะ (Constructs) ของแบบสอบทั้งสองฉบับ เป็นการตอบคำถามว่า “แบบสอบสองฉบับ นั้นวัดคุณลักษณะเดียวกันหรือไม่ ?” นั่นคือ พิจารณาถึงคะแนนจริงของแบบสอบ 2 ฉบับที่เกี่ยวข้องกัน สำหรับในบริบทของการเชื่อมโยงคะแนน อาจมีคุณลักษณะหรือโครงสร้างร่วมกันในระดับหนึ่ง แต่มิได้เป็นคุณลักษณะหรือโครงสร้างเดียวกัน
3. กลุ่มประชากรผู้สอบ (Populations) เป็นการตอบคำถามว่า “แบบสอบสองฉบับถูกออกแบบเพื่อวัดผู้สอบกลุ่มเดิมหรือไม่?”
4. เงื่อนไข/สถานการณ์ในการวัด (Measurement Conditions) เป็นการตอบคำถามว่า “แบบสอบสองฉบับมีเงื่อนไขในการวัดร่วมกันหรือไม่?” เงื่อนไขในการวัด เช่น ความยาวของแบบสอบ รูปแบบของแบบสอบ และการบริหารการสอบ ซึ่งในทฤษฎีการสรุปอ้างอิง (Generalizability Theory) เรียกเงื่อนไขในการวัดนี้ว่า Facets นั่นเอง อย่างไรก็ตามการกำหนดแบบแผนของแบบสอบก็เป็นส่วนหนึ่งของเงื่อนไขในการวัดโดยพิจารณาจากลักษณะความสำคัญ/ความร่วมกันของแบบสอบทั้ง 4 ประการที่กล่าวมาข้างต้น

สำหรับในกรณีของการเชื่อมโยงคะแนน Kolen & Brennan (2004) กล่าวว่าไม่ควรพิจารณาเฉพาะระดับความคล้ายคลึงกันของแบบสอบเท่านั้น แต่ควรพิจารณาถึงหลักเกณฑ์อื่นๆ ด้วย จากแนวคิดของ Feuer et al. (1999) ที่ได้เสนอเกี่ยวกับการวัดในกรณีที่แบบสอบขาดคุณลักษณะของความ

ร่วมกัน และจากการจำแนกประเภทตามแนวคิดของ Mislevy (1992) & Linn (1993) ทั้ง 2 แนวคิดใช้เกณฑ์ในการพิจารณาความร่วมกันของแบบสอบ คือ โครงร่างของแบบสอบ ที่มีลักษณะเดียวกัน และมีการกำหนดคุณลักษณะเฉพาะของแบบสอบที่มีลักษณะเดียวกัน โดยที่โครงร่างของแบบสอบคือแนวคิดในการตอบคำถามว่า แบบสอบนั้นมีโครงสร้างหรือคุณลักษณะที่คล้ายคลึงกันหรือไม่ ? ส่วนการกำหนดคุณลักษณะเฉพาะของแบบสอบ คือแนวคิดในการตอบคำถามว่า คุณลักษณะของแบบสอบทั้ง 2 ฉบับที่ต้องการวัดมีความคล้ายคลึงกันหรือไม่ ?

เมื่อนำแนวคิดของ Mislevy (1992) & Linn (1993), Feuer et al. (1999), Kolen & Brennan (2004) มาสังเคราะห์ร่วมกัน สามารถจำแนกประเภทของการเชื่อมโยงคะแนนในรูปแบบต่างๆ พร้อมทั้งหลักเกณฑ์ในการพิจารณา ดังตารางต่อไปนี้

ตารางที่ 2.3 ประเภทของการเชื่อมโยงคะแนนตามแนวคิดของ Mislevy (1992) & Linn (1993) และระดับของความคล้ายคลึงกันระหว่างแบบสอบสองฉบับ (อ้างใน Kolen & Brennan, 2004)

ประเภท	การสรุปอ้างอิง (Inferences)	โครงสร้าง/ คุณลักษณะ (Constructs)	กลุ่มประชากรผู้สอบ (Populations)	คุณลักษณะของแบบสอบที่ใช้ในการวัด (Measurement characteristics)
การปรับเทียบคะแนน (Equating)	มีลักษณะเดียวกัน	มีลักษณะเดียวกัน	มีลักษณะเดียวกัน	มีลักษณะเดียวกัน
การสร้างสเกลในแนวตั้ง (Vertical Scaling)	มีลักษณะเดียวกัน	มีลักษณะเดียวกัน/ คล้ายคลึงกัน	ไม่มีความคล้ายคลึงกัน	มีลักษณะเดียวกัน/ คล้ายคลึงกัน
คะแนนความสอดคล้อง (Concordance/Moderation)	มีลักษณะเดียวกัน	มีความคล้ายคลึงกัน	มีลักษณะเดียวกัน/ คล้ายคลึงกัน	มี/ไม่มีความคล้ายคลึงกัน
การฉายภาพ (Projection)	มี/ไม่มีความคล้ายคลึงกัน	มี/ไม่มีความคล้ายคลึงกัน	มีความคล้ายคลึงกัน	ไม่มีความคล้ายคลึงกัน
การปรับค่าทางสถิติ (Statistical Moderation)	มี/ไม่มีความคล้ายคลึงกัน	มี/ไม่มีความคล้ายคลึงกัน	มี/ไม่มีความคล้ายคลึงกัน	ไม่มีความคล้ายคลึงกัน

กล่าวโดยสรุป เมื่อพิจารณาถึงแนวคิดเกี่ยวกับการแบ่งประเภทของการเชื่อมโยงคะแนนและเกณฑ์ที่ใช้ในการแบ่ง ไม่ว่าจะเป็นแนวคิดของ Mislevy (1992) & Linn (1993) ที่มีการจำแนกรูปแบบการเชื่อมโยงคะแนน Feuer et al. (1999) ที่มีการศึกษาเกี่ยวกับการวัดในกรณีที่แบบสอบสองฉบับนั้นขาดความร่วมกัน/เกี่ยวข้องกัน จะสังเกตได้ว่าแต่ละแนวคิดยังคงไม่ใช่วิธีที่ดีที่สุดที่สามารถนำมาประยุกต์ใช้ในการเชื่อมโยงคะแนนได้ (Kolen & Brennan, 2004) ดังนั้นในการเชื่อมโยงคะแนนระหว่างแบบสอบ 2 ฉบับควรมีการกำหนดกรอบการดำเนินงานให้ชัดเจนและควรพิจารณา

ถึงการแปลผลของคะแนนที่ได้ด้วยว่า สามารถแสดงถึงความสามารถที่แท้จริงและคะแนนที่ได้เป็นตัวแทนที่ดีของผู้เรียนได้มากน้อยเพียงไร

ในรายละเอียดต่อไปจะกล่าวถึงแนวความคิดเกี่ยวกับการเชื่อมโยงคะแนนแบบพหุมิติ ซึ่งจะกล่าวถึง ที่มาและความสำคัญของการเชื่อมโยงคะแนนแบบพหุมิติ เทคนิคการเชื่อมโยงคะแนนตามโมเดล MIRT สำหรับข้อมูลที่มีการตรวจคะแนน 2 ค่า การหมุนแกนในตำแหน่งที่เหมาะสมและสเกลในตำแหน่งที่เหมาะสมการประมาณค่าการแปลงค่าพารามิเตอร์ตามโมเดล MIRT วิธีการเชื่อมโยงคะแนนตามโมเดล MIRT โครงสร้างมิติความสามารถ และเกณฑ์การเชื่อมโยงคะแนน โดยมีรายละเอียดดังต่อไปนี้

2.2 ที่มาและความสำคัญของการเชื่อมโยงคะแนนแบบพหุมิติ (Background of Multidimensional Linking Scores)

คำศัพท์ที่เกี่ยวข้องกับคำว่า “**การเชื่อมโยงคะแนนระหว่างแบบสอบ**” (Test Linking Scores) มีอยู่หลายคำ Mislevy (1992 อ้างใน Kolen & Brenan, 2004) & Linn (1993 อ้างใน Kolen & Brenan, 2004) ได้อธิบายให้เห็นความแตกต่างของศัพท์ทั้ง 4 คำ คือ การปรับเทียบคะแนน (Equating), การปรับคะแนนให้เป็นมาตรฐาน (Calibration/Vertical Scaling), การฉายภาพ (Projection) และการปรับค่าทางสถิติ (Statistical Moderation) ว่าต่างเป็นรูปแบบหนึ่งของการเชื่อมโยงคะแนน (Linking Scores) กระบวนการเชื่อมโยงคะแนนที่สำคัญ คือ การประเมินทั้งสองต้องมีเป้าหมายเหมือนกัน และมีเนื้อหาครอบคลุมลักษณะที่วัด โดย Mislevy & Linn ได้กล่าวถึง **การปรับเทียบคะแนน (Equating)** ว่าเป็นการเชื่อมโยงคะแนนจากแบบสอบต่างฉบับ แต่แบบสอบถูกสร้างด้วยลักษณะของความร่วมกันในด้านเนื้อหาและค่าสถิติระหว่างแบบสอบ 2 ฉบับ (Angoff, 1971) เมื่อพิจารณาเทียบกับ**การปรับทำคะแนนให้เป็นมาตรฐาน (Calibration)** หรือ **การสร้างสเกลแนวตั้ง (Vertical Scaling)** ซึ่งเป็นการเชื่อมโยงคะแนนจากแบบสอบ ที่มีโครงสร้างเดียวกันแต่มีระดับความยากที่แตกต่างกัน ถือได้ว่าเป็นการเชื่อมต่อที่มีเงื่อนไขน้อยกว่าการปรับเทียบคะแนน ส่วน**การฉายภาพ (Projection)** เป็นการเชื่อมโยงคะแนนจากแบบสอบที่มีโครงสร้างต่างกันด้วยวิธีการทำนาย และ **การปรับด้วยค่าสถิติ (Statistical Moderation)** เป็นการเชื่อมโยงคะแนนจากแบบสอบที่มีโครงสร้างต่างกัน มีตัวแปรปรับ (moderator) แฝงอยู่ในแบบสอบแต่ละฉบับ โดยทั้ง**การฉายภาพ (Projection)** และ **การปรับด้วยค่าสถิติ (Statistical Moderation)** เป็นการเชื่อมโยงคะแนนเมื่อข้อสอบวัดในโครงสร้างแตกต่างกัน

จากการพิจารณาคำทั้ง 4 คำข้างต้น จะเห็นว่ามีความแตกต่างของข้อตกลงเบื้องต้นที่กำหนด ซึ่งลดน้อยลงตามลำดับ โดยการปรับด้วยค่าสถิติ (Moderation) มีข้อตกลงที่ผ่อนปรนมากที่สุด ส่วนข้อกำหนดของการปรับเทียบคะแนน (Equating) จะมีข้อตกลงที่เข้มงวดมากที่สุด จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง พบว่าวิธีการปรับเทียบคะแนนของแบบสอบ (Test Equating) มีบทบาทสำคัญและเป็นเทคนิควิธีที่ได้รับการยอมรับตั้งแต่อดีตถึงปัจจุบัน อย่างไรก็ตามด้วยข้อจำกัดของความแกร่งต่อข้อตกลงเบื้องต้น และแบบสอบทั้งสองฉบับที่ทำการปรับเทียบต้อง

มุ่งวัดคุณลักษณะเดียวกัน ซึ่งเป็นการยากที่จะพบในสถานการณ์การสอบในปัจจุบัน ที่แบบสอบมีลักษณะการวัดหลายมิติ หรือหลายคุณลักษณะแฝง ส่งผลให้การรายงานผลของการปรับเทียบคะแนนในสเกลการวัดที่ต่างกันเกิดความไม่ถูกต้อง เมื่อใช้การปรับเทียบคะแนนตามโมเดล IRT ภายใต้ข้อตกลงเบื้องต้นที่สำคัญของความเป็นเอกมิติ (UIRT Equating) เมื่อแบบสอบที่สร้างขึ้นมุ่งวัดหลายมิติ ด้วยปัญหาของการปรับเทียบคะแนนที่ต้องอยู่บนฐานความคิดดังกล่าว จึงพัฒนาโมเดลการวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Item Response Theory; MIRT) เพื่อแก้ไขจุดบกพร่องของข้อตกลงเบื้องต้นดังกล่าว (Ackerman, 1992, 1994; Bock, Gibbons & Muraki, 1998; Reckase, 1985) ซึ่งโมเดล MIRT มีความเหมาะสมและสอดคล้องกับข้อมูลทั้งข้อมูลที่จำลองขึ้นมาและข้อมูลจริงที่มีหลายมิติมากกว่าโมเดลการวิเคราะห์แบบเอกมิติ หรืออีกนัยหนึ่งคือ โมเดลมีความสอดคล้องกับข้อมูลทำให้ผลการวิเคราะห์มีความถูกต้องและน่าเชื่อถือทั้งคุณสมบัติของความไม่แปรเปลี่ยนของค่าประมาณพารามิเตอร์ความสามารถ และความไม่แปรเปลี่ยนของค่าประมาณพารามิเตอร์ของข้อสอบ

อย่างไรก็ตามแม้จะมีการพัฒนาวิธีการปรับเทียบคะแนนแบบพหุมิติตามโมเดล MIRT (MIRT Equating) ตั้งแต่ปี ค.ศ. 1989 โดย Hirsch ต่อเนื่องมาจนกระทั่งในปี ค.ศ. 2000 (Li & Lissize, 2000) เริ่มมีการใช้ศัพท์คำว่า การเชื่อมโยงแบบพหุมิติ (MIRT Linking) แทนคำว่า การปรับเทียบคะแนน (MIRT Equating) และนิยมเขียนศัพท์คำว่า วิธีการเชื่อมโยงคะแนน (Linking Method/Approach) แทนคำว่าวิธีการปรับเทียบคะแนน (Equated Method) ต่อมาในปี ค.ศ. 2003 จนถึงปัจจุบันมีการประยุกต์ใช้วิธีการเชื่อมโยงคะแนนแบบพหุมิติตามโมเดล MIRT ในบริบทของการสร้างสเกลแนวตั้งแบบพหุมิติ (Vertical Scaling) แต่ยึดหลักวิธีการปรับเทียบแนวตั้งแบบพหุมิติ (Vertical Equating) เพื่อให้สอดคล้องกับสถานการณ์ของการวัดทางด้านการศึกษาและจิตวิทยาในสภาพจริงมากยิ่งขึ้น เช่น เนื้อหาของแบบสอบอาจมีโครงสร้างเหมือนกันหรืออาจคล้ายคลึงกันได้ตามบริบทที่ต้องการศึกษา (Min, 2003) โดยมีรายละเอียดของที่มาและความสำคัญเพื่อให้เห็นแนวโน้มของการเชื่อมโยงคะแนนแบบพหุมิติตั้งแต่อดีตจนถึงปัจจุบันดังต่อไปนี้

การเชื่อมโยงคะแนนระหว่างแบบสอบเริ่มมีขึ้นครั้งแรกในปี ค.ศ. 1989 โดย Hirsch ได้นำโมเดล MIRT มาใช้ในวิธีการเชื่อมโยงคะแนนที่เรียกว่า *การปรับเทียบคะแนนตามโมเดล MIRT (MIRT Equating)* เพื่อประโยชน์ในการนำวิธีการปรับเทียบคะแนนจากทฤษฎีสู่การปฏิบัติแต่วิธีการดังกล่าวมีความซับซ้อน และใช้ในกรณีศึกษากับข้อมูล 2 มิติ โดยอาศัยเมตริกซ์จากการหมุนแกน (Rotation Matrixes) ในการประมาณค่าพารามิเตอร์การเลื่อนแกน (Dilation) และประมาณค่าพารามิเตอร์ตำแหน่ง (Location) ต่อมา Oshima, Lee, & Davy (1997) ได้ศึกษาวิธีการปรับเทียบคะแนนที่ชัดเจนมากขึ้น โดยศึกษาจากวิธีการปรับเทียบคะแนนตามโมเดล MIRT 4 วิธี มีการประมาณค่าความสามารถจากเมตริกซ์การหมุนแกนและเวคเตอร์การแปลงคะแนน โดยใช้ฟังก์ชันทางคณิตศาสตร์ให้น้อยที่สุด เช่น การระบุความแตกต่างยกกำลังสองของพื้นผิวคุณลักษณะของแบบสอบทั้งสองฉบับ อย่างไรก็ตาม จากการศึกษาของ Oshima, Lee, & Davy

(1997) ไม่มีการประมาณค่าพารามิเตอร์การเลื่อนแกน (Dilation) โดยเมตริกซ์การหมุนแกน แต่ศึกษาจากรูปแบบความหลากหลายในการแปลงคะแนนรวม เพื่อนำมาใช้ในสถานการณ์เมตริกซ์การแปลงคะแนนแบบไม่ตั้งฉาก (Non-Orthogonal Transformation Matrix) ซึ่งต่อมาในปี ค.ศ. 2000 Oshima, Lee & Davy ได้นำผลงานวิจัยดังกล่าวเผยแพร่ลงวารสาร Journal of Education Measurement (JEM) เปลี่ยนศัพท์คำว่า “Multidimensional Equating” เป็น “Multidimensional Linking”

นอกจากนี้ Thompson et al. (1997) มีความพยายามที่จะพัฒนาวิธีการเชื่อมโยงคะแนนตามโมเดล MIRT (MIRT Linking) โดยศึกษาการเชื่อมโยงคะแนนที่ไม่ต้องใช้ผู้สอบร่วมหรือข้อสอบร่วมภายใต้สถานการณ์การสอบที่แบบสอบต่างกัน โดยใช้การสุ่มกลุ่มผู้สอบที่มีขนาดใหญ่ภายใต้ข้อตกลงเบื้องต้นว่า แกนของจุดกำเนิดและความสัมพันธ์ระหว่างแกนเป็นจุดร่วมกันของทั้งสองกลุ่ม โดยกลุ่มที่ได้ต้องเป็นกลุ่มสุ่มที่มีความเท่าเทียมกัน อย่างไรก็ตาม Reckase, Thompson, & Nering (1997) ได้กำหนดการหมุนแกนของแต่ละกลุ่มในระหว่างกระบวนการประมาณค่าพารามิเตอร์ ซึ่งมีความคล้ายคลึงกับการวิเคราะห์องค์ประกอบที่ไม่มีการกำหนดการหมุนแกนที่ตายตัว (Rotational Indeterminacy) วิธีการดังกล่าวจะนำไประบุกลุ่มของเนื้อหา (Cluster) โดยไม่ใช่กลุ่มของข้อสอบ โดยแบบสอบทั้งสองฉบับมีลักษณะแตกต่างกัน และจะมีการหมุนแกนจนกระทั่งอยู่ในสเกลเดียวกัน ต่อมา Li & Lissitz (2000) ได้ทำการประเมินประสิทธิภาพของการเชื่อมโยงคะแนนในเงื่อนไขที่ต่างกัน โดยอาศัยหลักของการปรับเทียบคะแนน

จะเห็นว่าในช่วงปี ค.ศ. 1997-2000 เริ่มมีผู้สนใจศึกษาการเชื่อมโยงคะแนน (Linking Score) ในบริบทของการปรับเทียบคะแนน แต่ไม่นิยมใช้ศัพท์คำว่าการปรับเทียบคะแนน (Equating) แต่ใช้คำว่าการเชื่อมโยงคะแนน (Linking) แทน และนำโมเดล MIRT มาใช้ศึกษากับข้อมูลที่มีลักษณะพหุมิติเพิ่มขึ้น พร้อมทั้งประเมินประสิทธิภาพของวิธีการเชื่อมโยงคะแนนเพื่อให้เห็นเป็นรูปธรรมมากขึ้น แต่ยังเป็นลักษณะเชิงทดลองโดยศึกษากับการจำลองข้อมูล และไม่สามารถระบุได้ชัดเจนว่าเงื่อนไขในการปรับเทียบใดดีที่สุดได้ ต่อมา Min (2003) ทำการศึกษาวิธีการเชื่อมโยงคะแนนในบริบทของการสร้างสเกลแนวตั้งแบบพหุมิติ (Multidimensional Vertical Scaling) แทนคำว่า การปรับเทียบคะแนนแนวตั้งแบบพหุมิติ (Multidimensional Vertical Equating) โดยศึกษาให้เห็นเป็นรูปธรรมมากขึ้น เริ่มมีการศึกษาจากข้อมูลจริงและนำไปประยุกต์ใช้กับการศึกษาพัฒนาการ มิติที่ศึกษามีจำนวนมากขึ้น

อย่างไรก็ตาม จากการศึกษาของ Min (2003) ได้ชี้ให้เห็นจุดอ่อนที่สำคัญของการเชื่อมโยงคะแนนแบบพหุมิติ คือ ถ้าจำนวนมิติที่ศึกษาไม่มาก วิธีดังกล่าวจะมีความถูกต้องในการปรับเทียบคะแนน แต่สำหรับกรณีที่โมเดลที่มีจำนวนมิติเพิ่มมากขึ้น การคำนวณจะจัดกระทำได้ยาก หรือไม่สามารถคำนวณได้ เนื่องจากปฏิสัมพันธ์ระหว่างคุณลักษณะเด่นของมิติข้ามชุดของแบบสอบ วิธีการหนึ่งที่สามารถจัดการกับปัญหานี้คือ การหมุนแกนแบบไม่ตั้งฉาก (Non-Orthogonal Procrustes Rototion) (Mulaik, 1972 อ้างใน Reckase & Martineau, 2004) ซึ่งจะทำการหมุนแกนเป็นไปโดยอัตโนมัติขึ้นอยู่กับความสอดคล้องที่ดีที่สุดของมิติของเมตริกซ์ปรับเทียบคะแนน (Equated Matrix)

กับมิติของเมตริกซ์พื้นฐาน (Based Matrix) ขณะเดียวกันการแปลงคะแนน โดยใช้การแปลงคะแนนแบบมุมแหลม (Oblique Transformation) จะช่วยลดความยุ่งยากในการคำนวณ โดยเฉพาะอย่างยิ่งการช่วยลดความจำเป็นหรือลดบทบาทของพารามิเตอร์การเลื่อนแกนที่เป็นสเกลลาหรือเมตริกซ์ โดยไม่นำมาคำนวณในสมการ ซึ่งแตกต่างจากวิธีของในอดีต

จากที่กล่าวมา วิธีของ Li & Lissitz (2000) & Min (2003) มีความเหมาะสมกับการศึกษาที่จำนวนมิติไม่มาก เช่น ศึกษา 2-3 มิติ เป็นต้น แต่เมื่อมีการศึกษาในจำนวนมิติมากขึ้น วิธีของ Reckase & Martineau (2004) จะเหมาะสมกว่า อย่างไรก็ตาม วิธีการข้างต้นไม่สามารถจำแนกมิติออกจากกันได้ การที่จะระบุความหมายของมิติตามแนวคิดของ MIRT เป็นเรื่องยากในการศึกษาจากข้อมูลเชิงประจักษ์ ซึ่งวิธีการโดยทั่วไปของการระบุหรือตีความหมายของมิติโดยการวิเคราะห์จัดกลุ่ม (Cluster Analysis) ดูเมตริกซ์ความสัมพันธ์ (Proximity Matrix) ของแกน ระหว่างเวคเตอร์อำนาจจำแนกของข้อสอบในพหุมิติ (Kim, 2001 อ้างใน Reckase & Martineau, 2004)

ต่อมา Reckase & Martineau (2006) ได้ศึกษาพัฒนาการภาคตัดขวางโดยการสร้างสเกลแนวตั้ง (Vertical Scaling) เน้นไปที่ความถูกต้องของมิติที่ศึกษา เนื่องจากแม้ว่าจะใช้แนวคิดของ MIRT มาช่วยในการวัดที่ประกอบไปด้วยหลายมิติ เพื่อแก้ปัญหาจาก UIRT แต่ไม่สามารถรับประกันได้ว่าวัดได้ตรงกับพัฒนาการของนักเรียนหรือไม่ ซึ่งผลกระทบที่แท้จริงอาจจะมาจากการสอนก็เป็นได้ และประเด็นเกี่ยวกับความตรงและความเที่ยง ซึ่งการกำหนดมิติที่เหมาะสมเป็นเรื่องยากและยังขาดความชัดเจน

จากการศึกษาของ Oshima et al. (2000) Li & Lissize (2000) Min (2003) และ Reckase & Martineau (2004, 2006) ถือว่าเป็นช่วงเวลาสำคัญในการพัฒนาวิธีการเชื่อมโยงคะแนนที่เห็นผลในทางปฏิบัติ และมีประโยชน์ในการประยุกต์ใช้ในสถานการณ์จริงมากขึ้น แต่ยังขาดความชัดเจนเกี่ยวกับวิธีการใดเหมาะสมกับสถานการณ์ใด ซึ่งจากการศึกษาในอดีต ยังไม่มีการศึกษาเปรียบเทียบในวิธีที่พัฒนาขึ้นอย่างจริงจัง จนกระทั่ง Yon (2006) ต้องการตอบคำถามดังกล่าวจึงทำการเปรียบเทียบและประเมินคุณภาพของวิธีการเชื่อมโยงคะแนน 2 วิธี คือ วิธีฟังก์ชันลักษณะข้อสอบ (Test Characteristic Function; TCF) (Oshima et al., 2000) และ วิธี Non-Orthogonal Procrustes (NOP) (Reckase & Martineau, 2004) โดยศึกษาจากการจำลองข้อมูลและข้อมูลจริงเพื่อนำไปประยุกต์ใช้ในการสร้างสเกลแนวตั้งแบบพหุมิติ (Multidimensional Vertical Scaling) ในการศึกษาพัฒนาการของผู้สอบทั้งหมด 3 ระดับชั้นที่แตกต่างกัน ซึ่งต่อมา Yon (2007) มุ่งศึกษาถึงปัญหาความถูกต้องของเมตริกซ์ในการหมุนแกนเมื่อเกิดการเปลี่ยนแปลงโครงสร้าง (Construct Shift) จากสถานการณ์ของการจำลองข้อมูล โดยกำหนดให้แบบสอบมีจุดเน้นที่ต่างกัน โดยพิจารณาความแตกต่างระหว่างแบบสอบฐานกับแบบสอบปรับเทียบคะแนน ในการประมาณค่าพารามิเตอร์จากการจำลองข้อมูลใน 2 มิติ กับผู้สอบที่แตกต่างกัน 2 ระดับชั้น ซึ่งถือได้ว่างานวิจัยของ Yon (2006, 2007) เป็นงานวิจัยที่มีความทันสมัยมากที่สุดในปัจจุบัน ที่ทำการศึกษาเกี่ยวกับวิธีการเชื่อมโยงคะแนนแบบพหุมิติตามโมเดล MIRT เพื่อต้องการนำไปสู่คำตอบของการเลือกวิธีการปรับเทียบคะแนนให้เหมาะสมกับสถานการณ์ที่ต้องการศึกษา

กล่าวโดยสรุป การเชื่อมโยงคะแนนตามโมเดล MIRT มีการพัฒนาอย่างต่อเนื่อง โดยศึกษาในรูปแบบของการปรับเทียบคะแนน (Equating) เพื่อให้ทันกับยุคสมัยและความก้าวหน้าในการศึกษาค้นคว้า โดยในระยะเวลา 19 ปี ที่ผ่านมา การศึกษาการเชื่อมโยงคะแนนแบบพหุมิติ (MIRT Linking) ในช่วงแรก (1989-1995) ยังเป็นการศึกษาในเชิงทดลองและไม่สามารถนำไปปฏิบัติได้จริง และต้องอาศัยฟังก์ชันทางคณิตศาสตร์ที่ซับซ้อน ต่อมาในช่วงปี ค.ศ. 1996-2000 นักวัดผลเริ่มตื่นตัวในการศึกษามากขึ้น โดยเฉพาะการประเมินประสิทธิภาพของการเชื่อมโยงคะแนน เนื่องจากเล็งเห็นถึงความสำคัญที่สามารถนำมาแก้จุดอ่อนของการเชื่อมโยงคะแนนแบบเอกมิติ (UIRT Linking) ได้ สำหรับในช่วงปี ค.ศ. 2001-ปัจจุบัน พบว่าเริ่มนำโมเดล MIRT มาประยุกต์ใช้ปรับเทียบสเกลแนวตั้งแบบพหุมิติ (Multidimensional Vertical Equating) เพื่อศึกษาพัฒนาการของผู้เรียน โดยเป็นการศึกษาภาคตัดขวาง เพื่อประโยชน์ในการพัฒนาระบบการศึกษาในประเทศ

ข้อสังเกตประการสำคัญคือ แม้จะมีการศึกษาวิธีการเชื่อมโยงคะแนนในรูปแบบของการปรับเทียบคะแนน แต่ไม่นิยมใช้คำว่า MIRT Equating แต่ใช้คำว่า MIRT Linking ดังนั้นเพื่อให้การใช้ศัพท์สอดคล้องกับงานวิจัยที่ผ่านมาและมีความสากล ผู้วิจัยจึงใช้ศัพท์คำว่า **“การเชื่อมโยงคะแนน” (Linking)** แทนคำว่า **“การปรับเทียบคะแนน” (Equating)** แต่ยังเป็นการศึกษาในบริบทของการปรับเทียบคะแนน

จากที่กล่าวมาจะเห็นว่า การเชื่อมโยงคะแนนให้ประสบความสำเร็จขึ้นอยู่กับความถูกต้องและเหมาะสมของเงื่อนไขที่ใช้ในสถานการณ์ต่าง ๆ ดังนั้นในรายละเอียดต่อไปผู้วิจัยขอนำเสนอเทคนิคการเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติสำหรับข้อมูลที่มีการตรวจคะแนนแบบ 2 ค่า เพื่อเป็นการการขยายองค์ความรู้เกี่ยวกับการเชื่อมโยงคะแนนแบบพหุมิติ (MIRT Equating) ในการนำไปสู่การประยุกต์ใช้ได้อย่างถูกต้อง แม่นยำ และมีความยุติธรรม

2.3 เทคนิคการเชื่อมโยงคะแนนตามโมเดลการตอบสนองข้อสอบแบบพหุมิติสำหรับข้อมูลที่มีการตรวจคะแนนแบบ 2 ค่า (MIRT Linking Methods for Dichotomous Item Response)

การเชื่อมโยงคะแนนตามโมเดล MIRT ตั้งแต่อดีตถึงปัจจุบัน ได้ศึกษาเฉพาะบริบทของการปรับเทียบคะแนน (MIRT Equating) และการสร้างสเกลแนวตั้งแบบพหุมิติ (Multidimensional Vertical Scaling) ดังที่กล่าวในรายละเอียดข้างต้น ซึ่งโดยทั่วไปจะนิยมประมาณค่าพารามิเตอร์โดยใช้โมเดลโลจิสติกแบบพหุมิติ มีลักษณะคล้ายคลึงกับโมเดลโลจิสติกแบบเอกมิติ ในที่นี้ผู้วิจัยจะนำเสนอเฉพาะเทคนิคการเชื่อมโยงคะแนนตามตามโมเดล MIRT สำหรับข้อมูลที่มีการตรวจคะแนนแบบ 2 ค่า เทคนิควิธีของการประมาณค่าการแปลงค่าพารามิเตอร์ตามโมเดล MIRT การกระจายความคลาดเคลื่อนสำหรับการประมาณค่าพารามิเตอร์ วิธีการเชื่อมโยงคะแนน และโครงสร้างมิติความสามารถ โดยมีรายละเอียดดังนี้

2.3.1 โมเดลโลจิสติกตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (MIRT Model)

โมเดล MIRT ได้รับการพัฒนาขึ้นเพื่อใช้อธิบายผลของความสามารถที่มีหลายมิติของแบบสอบที่นำมาศึกษา โดยมีข้อตกลงเบื้องต้นว่า *ประชากรมีความหลากหลายของทักษะและความสามารถ* ซึ่งใช้โมเดลเพื่อทำนายโอกาสในการตอบข้อสอบได้ถูกต้อง เนื่องจากโมเดล MIRT เป็นเทคนิคพิเศษของการวิเคราะห์องค์ประกอบ หรือโมเดลสมการเชิงโครงสร้าง ซึ่งแผ่ขยายมาจากโมเดล UIRT (Reckase, 1997) โดยในช่วงปี 1970 ถึงช่วงต้นปี 1980 เป็นต้นมา มีนักวิจัยเป็นจำนวนมากที่ให้ความสนใจกับการประยุกต์ใช้โมเดล MIRT (Reckase, 1972; Mulaik, 1972; Sympson, 1978; Whitely, 1980; Mckinly & Reskase, 1982)

เมื่อพิจารณาในบริบทของข้อตกลงเบื้องต้นที่แสดงความสัมพันธ์ระหว่างมิติความสามารถ (Ability Dimension) ของโมเดล MIRT พบว่า สามารถแบ่งได้เป็น 2 ประเภท คือ 1) โมเดลที่การประมาณค่าความสามารถ (θ) ในมิติต่างกันไม่สามารถชดเชยกันได้ (Non-Compensatory Model) และ 2) โมเดลที่การประมาณค่าความสามารถ (θ) ในมิติต่างกันสามารถชดเชยกันได้ (Compensatory Model) ซึ่งโมเดลทั้งสองมีความแตกต่างกันของความน่าจะเป็นในการตอบข้อสอบของบุคคลได้ถูกต้อง จะเกี่ยวข้องกับความสัมพันธ์ระหว่างมิติความสามารถ โดยมีรายละเอียด ดังนี้

(1) โมเดลที่การประมาณค่าความสามารถ (θ) ในมิติต่างกันไม่สามารถชดเชยกันได้ (Non-Compensatory Model) (Embretson, 1984; Sympson, 1978) เป็นโมเดลที่ในแต่ละมิติ มีค่าพารามิเตอร์ความยากและอำนาจจำแนกเฉพาะของแต่ละมิติแยกจากกัน ซึ่งสมการที่ได้จะแสดงความสัมพันธ์กับมิติที่ต่างกัน สามารถคำนวณได้จากการคูณความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องของแต่ละมิติ เมื่อพิจารณาสมการของโมเดลมีความหมายว่า การที่มีความสามารถต่ำของคุณลักษณะแฝงในมิติหนึ่ง อาจจะสามารถทดแทนความน่าจะเป็นในการตอบข้อสอบถูกต้องได้น้อย หรือทดแทนได้เพียงบางส่วนจากคุณลักษณะแฝงที่มีความสามารถสูงในอีกมิติหนึ่ง โดยมีความน่าจะเป็นน้อยที่สุดสำหรับชุดข้อสอบที่สามารถชดเชยได้ เป็นช่วงขีดจำกัดบนของโอกาสในการตอบข้อสอบได้ถูกต้อง บางครั้งเรียกโมเดลนี้ว่า **โมเดลที่การประมาณค่าความสามารถ (θ) ในมิติต่างกันสามารถชดเชยได้บางส่วน (Partially Compensatory Model)** เนื่องจากจะเพิ่มความสามารถใน θ ในมิติที่หนึ่ง ส่งผลได้น้อยต่อมิติอื่นๆ เมื่อ θ_j เป็นเวคเตอร์ความสามารถของบุคคลที่ j (Reckase, 1997) สามารถแสดงสมการของโมเดลแบบพหุมิติชนิด 3 พารามิเตอร์ตามโมเดลดังกล่าว (Multidimensional Three-Parameter Non-Compensatory Model) ที่เสนอโดย Sympson (1978) ดังต่อไปนี้

$$P\left(x_{is}=1\middle|a_i,d_i,\theta_i,c_i\right)=c_i+(1-c_i)\prod_{k=1}^{m}\frac{\exp\left(a_{ik}^{'}\theta_{jk}+d_{ik}\right)}{1+\exp\left(a_{ik}^{'}\theta_{jk}+d_{ik}\right)} \quad \text{.....(2.19)}$$

เมื่อ $P(X_{ij}=1|a_i,d_i,\theta_i,c_i)$ แทน ความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง (x_{ij} = 1) ของข้อสอบข้อที่ i ของบุคคลที่ j ในมิติที่ m

โดย X_{ij} แทน คะแนนที่ได้ 0 (ตอบผิด) หรือ 1 (ตอบถูก) ในข้อสอบข้อที่ i ของบุคคลที่ j

θ_j แทน ความสามารถแฝงของเวคเตอร์ m มิติ (m-dimensional vector)

a_i แทน ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบในเวคเตอร์ m มิติ

d_i แทน ค่าพารามิเตอร์ความยากของข้อสอบในเวคเตอร์ m มิติ

c_i แทน ค่าพารามิเตอร์โอกาสการเดาข้อสอบ

k แทน การศึกษาในมิติที่ k

θ_{jk} แทน ค่าของเวคเตอร์ของคุณลักษณะแฝงในมิติที่ k

a_{ik} แทน ค่าของเวคเตอร์อำนาจจำแนกในมิติที่ k

d_{ik} แทน ค่าเวคเตอร์ความยากในมิติที่ k

(2) โมเดลที่การประมาณค่าความสามารถ (θ) ในมิติต่างกันสามารถชดเชยกันได้ (Compensatory Model) ได้มีผู้ศึกษาไว้หลายกลุ่ม (Lord & Novick, 1968; McDonald, 1967; Reckase, 1985; 1995) เป็นโมเดลที่มีลักษณะทางบวก (Additive in the Logit) เช่น สำหรับผู้ที่มีความสามารถต่ำในมิติที่หนึ่งสามารถชดเชยความสามารถจากความการมีความสามารถสูงในอีกมิติหนึ่งนั่นเอง สามารถแสดงสมการของโมเดลแบบพหุมิติชนิด 3 พารามิเตอร์ตามโมเดลดังกล่าว (Multidimensional Three-Parameter Compensatory Model) ที่เสนอโดย McKinley & Reckase (1983) & Reckase (1985, 1995) ดังต่อไปนี้

$$P\left(U_{ij}=1\middle|a_i,d_i,\theta_i,c_i\right)=c_i+(1-c_i)\frac{\exp(a_i'\theta_j+d_i)}{1+\exp(a_i'\theta_j+d_i)} \quad \text{.....(2.20)}$$

เมื่อ $P(U_{ij}=1|a_i,d_i,\theta_i,c_i)$ แทน ความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง (x_{ij} = 1) ของข้อสอบข้อที่ i ของบุคคลที่ j ในมิติที่ m

โดย X_{ij} แทน คะแนนที่ได้ 0 (ตอบผิด) หรือ 1 (ตอบถูก) ในข้อสอบข้อที่ I ของบุคคลที่ j

θ_j แทน ความสามารถแฝงของเวคเตอร์ m มิติ (m-dimensional vector)

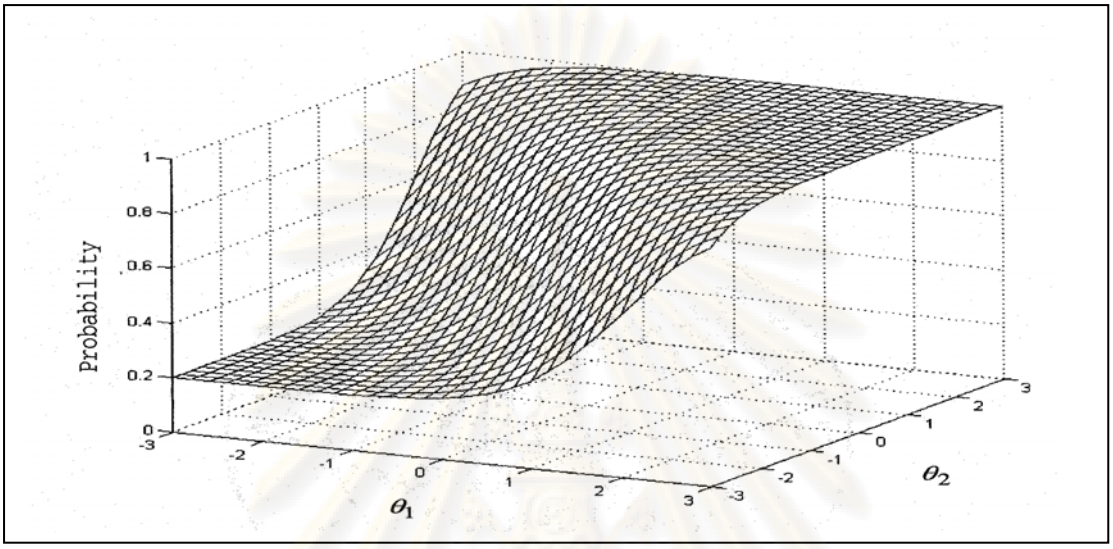
a_i แทน เวคเตอร์ของค่าพารามิเตอร์ที่มีความสัมพันธ์กับอำนาจจำแนกของข้อออบข้อที่ i ในแต่ละมิติของ m

d_i แทน ค่าพารามิเตอร์ที่มีลักษณะเป็นสเกลลาที่มีความสัมพันธ์กับความยากของข้อสอบข้อที่ i

c_i แทน ค่าพารามิเตอร์โอกาสการเดา สำหรับคนที่มีความสามารถต่ำจะตอบข้อสอบได้ถูกต้องใน m มิติ

โมเดลจะแสดงถึงโอกาสในการตอบข้อสอบได้ถูกต้องซึ่งมีลักษณะของการเพิ่มขึ้นทางเดียวอยู่ในช่วง 0-1 ดังตัวอย่างที่แสดงในแผนภาพที่ 2.6 แสดงพื้นผิวการตอบสนองข้อสอบแบบ 3

พารามิเตอร์ใน 2 มิติ มีลักษณะที่คล้ายกับโมเดล UIRT ประกอบด้วย a_1=1, a_2=1, c=0.2 และ d=1 ซึ่งความสูงแสดงถึงความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องในมิติความสามารถ (θ_1, θ_2) โดยความน่าจะอยู่เป็นในช่วง 0.2-1 เนื่องจากมีค่าโอกาสการเดาเท่ากับ 0.2 ข้อสอบข้อนี้จะวัดใน 2 มิติ เพราะมีค่าอำนาจจำแนกของข้อสอบทั้ง 2 มิติ ดังแผนภาพ



แผนภาพที่ 2.6 พื้นผิวการตอบข้อสอบ ซึ่งมี a_1=1, a_2=1, c=0.2 และ d=1

จากสมการ (2.20) จะเห็นว่าโมเดล MIRT มีลักษณะคล้ายกับโมเดล UIRT ซึ่งโมเดลโลจิสติกแบบพหุมิติ มีลักษณะใกล้เคียงกับโมเดลปกติสะสมแบบพหุมิติ ชนิด 3 พารามิเตอร์ (Knol & Berger, 1988) อย่างไรก็ตามเมื่อเปรียบเทียบสมการจากทั้งสองโมเดลระหว่างโมเดลโลจิสติกกับโมเดลปกติสะสม (Normal Ogive Model) โมเดลโลจิสติกเป็นที่นิยมใช้เนื่องจากมีความสลับซับซ้อนน้อยกว่า (Embretson & Reise, 2000) สาเหตุที่โมเดลแบบปกติสะสมมีความซับซ้อนมากกว่าเนื่องจาก ต้องมีการอินทิเกรตลักษณะการกระจาย ซึ่งสามารถแสดงได้ดังสมการต่อไปนี้

$$P\left(x_{is}=1\middle|\theta_j,\beta_i,\alpha_i,\gamma_i\right)=\gamma_i+(1-\gamma_i)\int_{-\infty}^{\alpha_i(\theta_j-\beta_i)}\frac{1}{(2\pi)^{1/2}}\exp\left(-t^2/2\right)dt \quad \text{.....(2.21)}$$

เมื่อ $P(X_{ij}=1|a_i,d_i,\theta_i,c_i)$ แทน ความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง (x_{ij} = 1) ของข้อสอบข้อที่ i ของบุคคลที่ j ในมิติที่ m

โดย X_{ij} แทน คะแนนที่ได้ 0 (ตอบผิด) หรือ 1 (ตอบถูก) ในข้อสอบข้อที่ i ของบุคคลที่ j

θ_j แทน ค่าพารามิเตอร์ที่อธิบายความสามารถของบุคคลที่ j

α_i แทน ค่าพารามิเตอร์ที่มีความสัมพันธ์กับอำนาจจำแนกของข้อสอบข้อที่ i ในแต่ละมิติของ m

β_i แทน ค่าพารามิเตอร์ที่มีลักษณะเป็นสเกลลาที่มีความสัมพันธ์กับความยากของข้อสอบข้อที่ i

โดยทั่วไป โมเดลที่นิยมมากที่สุดในการศึกษาการปรับเทียบคะแนนตามโมเดล MIRT คือ โมเดลที่การประมาณค่าความสามารถ (θ) ในมิติต่างกันสามารถชดเชยกันได้ (Compensatory Model) เนื่องจากโมเดลที่การประมาณค่าความสามารถ (θ) ในมิติต่างกันไม่สามารถชดเชยกันได้ (Non-Compensatory Model) มีความยุ่งยากและซับซ้อนสำหรับการประมาณค่าและมีความยุ่งยากในทางปฏิบัติ อีกทั้งขั้นตอนในการคำนวณไม่มีประสิทธิภาพเพียงพอในการประมาณค่าพารามิเตอร์ (Knol & Berger, 1988) แต่เมื่อพิจารณาความสอดคล้องของโมเดลทั้งสองชนิดไม่สามารถจำแนกความแตกต่างในทางปฏิบัติได้ (Spray, Davey, Reckase, Ackerman, & Carlson, 1990)

เมื่อพิจารณาโมเดล MIRT ที่ต่างจาก UIRT มีหลายประการที่สำคัญ คือ ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบและค่าพารามิเตอร์ความสามารถของผู้สอบในโมเดล MIRT อยู่ในรูปของเวคเตอร์มากกว่าสเกลลา และค่าพารามิเตอร์ที่เกี่ยวข้องกับความยากของข้อสอบ (Difficulty-Related Parameter) คือค่าการรวมกันของค่าความยากและค่าอำนาจจำแนกของข้อสอบในแต่ละมิติ โดยค่าอำนาจจำแนกมีความสัมพันธ์ว่ามีความไว (Sensitive) ของความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องเป็นอย่างไร ในขณะที่โมเดล UIRT ค่าพารามิเตอร์อำนาจจำแนกเป็นสัดส่วนของความชันในโค้ง ICC ณ จุดเปลี่ยนโค้งที่มีความชันสูงสุด (Steepest Slope)

ค่าอำนาจจำแนก (MDISC)

เมื่อเปรียบเทียบค่าอำนาจจำแนก ตามโมเดล MIRT ค่าอำนาจจำแนก คือ มีลักษณะเป็นเวคเตอร์ (a-เวคเตอร์) แต่สามารถแปลความหมายในทิศทางเดียวกันกับค่าพารามิเตอร์ในโมเดล UIRT (Lord, 1980) ตามด้วยส่วนประกอบของเวคเตอร์ที่มีความสัมพันธ์กับความชันของพื้นผิวการตอบสนองข้อสอบในตำแหน่งที่สอดคล้องกับแกนความสามารถ (θ-Axis) โดยขึ้นอยู่กับมุมระหว่าง เวคเตอร์กับแกนที่ต้องการหาคำตอบ ซึ่งค่าอำนาจจำแนกของข้อสอบตามโมเดล MIRT (Discriminating Power of Multidimensional Item) (Reckase, 1985; 1995; Reckase & Mckinley, 1991) สามารถแสดงดังสมการ

$$MDISC_i = \sqrt{\sum_{k=1}^{M} a_{ik}^2} \quad \text{.....(2.22)}$$

โดย $MDISC_i$ แทน ค่าอำนาจจำแนกรวมของข้อสอบข้อที่ i ใน m มิติ

M แทน จำนวนของมิติใน θ-Space

a_{ik} แทน ส่วนประกอบของ a_i-เวคเตอร์ของมิติที่ k

จากสมการจะเห็นได้ว่า $MDISC_i$ เป็นสัดส่วนของความชันพื้นผิวการตอบสนองข้อสอบตรงตำแหน่งที่มีความชันสูงสุด จึงสังเกตได้ว่ามีลักษณะที่คล้ายกับค่าพารามิเตอร์อำนาจจำแนกในโมเดล UIRT (Carlson, 1987; Reckase & McKinley, 1991)

ค่าตำแหน่งของข้อสอบ (Location of an Item)

ส่วนประกอบที่สำคัญประการหนึ่งของโมเดล MIRT คือ ตำแหน่งของข้อสอบในมิติความสามารถ (θ-Space) สามารถแทนด้วยเวคเตอร์ความยาวที่สอดคล้องกับค่า $MDISC_i$ ซึ่ง

ตำแหน่งของเวคเตอร์ คำนวณได้จากการพิจารณาตำแหน่งของ Cosines ซึ่งสามารถแสดงได้ดังสมการ

$$Cos\alpha_{ik} = \frac{\alpha_{ik}}{\sqrt{\sum_{k=1}^{M} a_{ik}^2}} \quad \text{.....(2.23)}$$

โดย M แทน จำนวนของมิติใน θ-Space

a_{ik} แทน ส่วนประกอบของ a_i-เวคเตอร์ของมิติที่ k

ค่าความยากของข้อสอบ ($MDIFF_i$)

Reckase (1985) ได้ให้ความหมายของค่าความยากของข้อสอบตามโมเดล MIRT ($MDIFF_i$) โดยเป็นฟังก์ชันที่ประกอบด้วยทิศทางใน Space ที่ได้แสดงในสมการ 2.24 และระยะทางจากจุดกำเนิดไปยังจุดที่มีความชันสูงสุด (Steepest Slope) ดังแสดงในสมการต่อไปนี้

$$MDIFF_i = \frac{-d_i}{MDISC_i} \quad \text{.....(2.24)}$$

โดย d_i แทน พารามิเตอร์ที่มีความสัมพันธ์กับค่าความยากของข้อสอบในแบบสอบ

จากสมการที่ 2.24 พบว่าค่า d_i เป็นพารามิเตอร์ที่มีความสัมพันธ์กับค่าความยากของข้อสอบในแบบสอบ อย่างไรก็ตาม $MDIFF_i$ ไม่สามารถแปลความหมายเหมือนค่าความยากของข้อสอบในโมเดล UIRT ซึ่งค่าของ $MDIFF_i$ บ่งชี้ถึงระยะทางจากจุดกำเนิดของ θ-Space ไปยังจุดที่มีความชันสูงสุด (Steepest Slope)

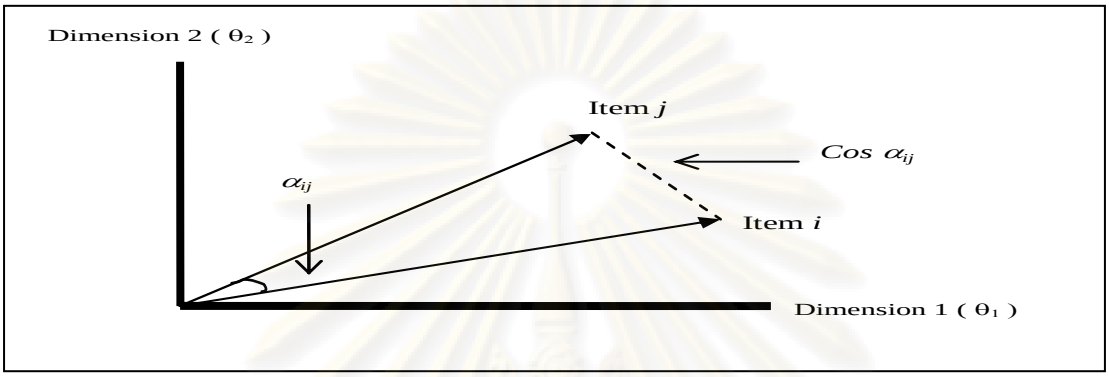
ตำแหน่งของค่าความยากและค่าอำนาจจำแนกแบบพหุมิติ (α_{ik} หรือ $\cos\alpha_{ik}$)

ตำแหน่งของค่าความยากและค่าอำนาจจำแนกของมิติใน Space สามารถแสดงได้ดังสมการที่ 2.25

$$\alpha_{ik} = \arccos\frac{\alpha_{ik}}{MDISC_i} \text{ หรือ } \cos\alpha_{ik} = \frac{\alpha_{ik}}{MDISC_i} \quad \text{.....(2.25)}$$

Reckase, Ackerman, & Carlson (1988) พบว่าตำแหน่งเวคเตอร์ของข้อสอบเป็นตัวบ่งชี้น้ำหนักโดยรวมของการวัดความสามารถที่ดีที่สุด ดังนั้นชุดข้อสอบมีตำแหน่งมุมหรือค่า Cosines ที่เหมือนกันหรือใกล้เคียงกัน เมื่อมีความแตกต่างที่มากกว่า 1 ในชุดของข้อสอบกับความคล้ายคลึงกันของตำแหน่ง Cosines ในมิติแฝง ซึ่งแต่ละชุดสามารถรวมความแตกต่างของแต่ละมิติความสามารถใน Space ที่แผ่ขยายระหว่างเวคเตอร์ นั่นคือ มีการปรับชุดที่ส่งผลต่อระดับของความเป็นพหุมิติในชุดที่ศึกษา ดังนั้นการประเมินตำแหน่งของมุมและค่า Cosines ของเวคเตอร์ข้อสอบในการวิเคราะห์ MIRT ให้ความหมายของการระบุเป็นเซต นั่นคือเซตที่มีการวัดความสามารถเดียวกัน และประโยชน์ต่อการพิจารณาโครงสร้างของแบบสอบ (Miller & Hirsch, 1992)

ระยะทางของมุมระหว่างชุดของแบบสอบ 2 ชุด ดังแสดงในแผนภาพที่ 2.7 ซึ่งจากการชี้ของลูกศรบ่งชี้ถึงตำแหน่งของความชันสูงสุด (Steepest Slope) ของพื้นผิวการตอบข้อสอบสำหรับข้อที่ i และ j ไปยังระนาบ ในมิติของ θ_1 และ θ_2 ถ้า cos a_{ij} มีค่าเข้าใกล้ 0 แสดงว่าข้อที่ i และ j วัดคุณลักษณะแฝงที่แตกต่างกัน ในขณะเดียวกัน ถ้า cos a_{ij} มีค่าเข้าใกล้ 1 แสดงว่าข้อที่ i และ j วัดคุณลักษณะแฝงเดียวกันหรือมิติเดียวกัน



แผนภาพที่ 2.7 ระยะทางของมุมระหว่างมิติที่ 1 และมิติที่ 2

ค่าโอกาสการเดาของข้อสอบ (c_i) มีความหมายเดียวกันทั้งโมเดล UIRT และ MIRT ซึ่งค่าพารามิเตอร์ c_i เป็นความน่าเป็นของผู้สอบที่มีความสามารถต่ำจะตอบข้อสอบได้ถูกต้องในทุกมิติในโมเดล

จากการพิจารณาการประมาณค่าพารามิเตอร์ตามโมเดล MIRT สามารถสรุปความแตกต่างของการประมาณค่าพารามิเตอร์ระหว่างโมเดล UIRT กับโมเดล MIRT ได้ดังตารางต่อไปนี้

ตารางที่ 2.4 ความแตกต่างของการประมาณค่าพารามิเตอร์ระหว่างโมเดล UIRT กับโมเดล MIRT

ค่าพารามิเตอร์	โมเดล UIRT	โมเดล MIRT
ความน่าจะเป็นของการตอบข้อสอบได้ถูกต้อง $P_i(\theta)$	$P(X_{ij}=1 \mid a_i,b_i,\theta_i,c_i)$ แทน ความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง (x_{ij} = 1) ของข้อสอบข้อที่ i ของบุคคลที่ j ใน 1 มิติ	$P(X_{ijk}=1 \mid a_i,d_i,\theta_i,c_i)$ แทน ความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง (x_{ij} = 1) ของข้อสอบข้อที่ i ของบุคคลที่ j ในมิตที่ k ซึ่งมีทั้งหมด m มิติ
ค่าความสามารถ (θ)	อยู่ในรูปของสเกลลา	อยู่ในรูปของเวคเตอร์
ค่าอำนาจจำแนกของข้อสอบ	-อยู่ในรูปของสเกลลา - ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ (a) เป็นสัดส่วนของความชันในโค้ง ICC ณ จุดเปลี่ยนโค้งที่มีความชันสูงสุด (Steepest Slope)	-อยู่ในรูปของเวคเตอร์ (a-เวคเตอร์)ที่มีความสัมพันธ์กับความชันของพื้นผิวการตอบข้อสอบ θ-Axis ขึ้นอยู่กับมุมระหว่าง a-เวคเตอร์กับแกนที่ต้องการหาคำตอบ - ค่าอำนาจจำแนกรวมของข้อสอบข้อที่ i ใน m มิติ ($MDISC_i$) เป็นสัดส่วนของความชันพื้นผิวการตอบสนองข้อสอบที่จุดที่มีความชัดสูงสุด (Steepest Slope)

ค่าพารามิเตอร์	โมเดล UIRT	โมเดล MIRT
ค่าความยากของข้อสอบ	-อยู่ในรูปของสเกลลา -ค่าพารามิเตอร์ความยากของข้อสอบ (b) เป็นตำแหน่งของโค้งบนสเกลของความสามารถ (θ) ที่ทำให้มีโอกาสตอบข้อสอบได้ถูกต้องเท่ากับ $\frac{1+c_i}{2}$	-อยู่ในรูปของเวคเตอร์ -$MDIFF_i$ เป็นฟังก์ชันที่ประกอบด้วยทิศทางใน Space (d_i) และระยะทางจากจุดกำเนิดไปยังจุดที่มีความชันสูงสุด ($MDISC_i$) -ค่าของ $MDIFF_i$ บ่งชี้ถึงระยะทางจากจุดกำเนิดของ θ-Space ไปยังจุดที่มีความชันสูงสุด (Steepest Slope)
ค่าตำแหน่งของข้อสอบ	-	เป็นเวคเตอร์ความยาวที่สอดคล้องกับค่า $MDISC_i$ ซึ่งตำแหน่งของเวคเตอร์สามารถคำนวณได้จากการพิจารณาจากตำแหน่งของ Cosines
พารามิเตอร์ที่มีความสัมพันธ์กับค่าความยากของข้อสอบ	-	ค่าพารามิเตอร์ที่เกี่ยวข้องกับความยากของข้อสอบ (Difficulty-Related Parameter) หรือ d_i คือค่าการรวมกันของค่าความยากของข้อสอบและค่าอำนาจจำแนกในแต่ละมิติ

กล่าวโดยสรุป โมเดล MIRT ที่นำมาใช้ในการเชื่อมโยงคะแนนมี 2 ประเภทที่สำคัญ เมื่อพิจารณาในบริบทของข้อตกลงเบื้องต้นที่แสดงความสัมพันธ์ระหว่างมิติความสามารถ (Ability Dimension) คือ 1) โมเดลที่การประมาณค่าความสามารถ (θ) ในมิติต่างกันไม่สามารถชดเชยกันได้ (Non-Compensatory Model) และ 2) โมเดลที่การประมาณค่าความสามารถ (θ) ในมิติต่างกันสามารถชดเชยกันได้ (Compensatory Model) ซึ่งโมเดลทั้งสองมีความแตกต่างกันของความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องในแต่ละข้อ เกี่ยวข้องกับความสัมพันธ์ระหว่างมิติความสามารถ ซึ่งโดยทั่วไป โมเดลที่นิยมมากที่สุดในการศึกษาการเชื่อมโยงคะแนน คือโมเดลที่การประมาณค่าความสามารถ (θ) ในมิติต่างกันสามารถชดเชยกันได้ (Compensatory Model) เนื่องจากไม่ยุ่งยากและซับซ้อนสำหรับการประมาณค่าตามโมเดลและมีความยืดหยุ่นในทางปฏิบัติ ซึ่งขั้นตอนในการคำนวณมีประสิทธิภาพที่เพียงพอในการประมาณค่าพารามิเตอร์

เมื่อพิจารณาโมเดล MIRT ที่ต่างจาก UIRT ประการสำคัญ คือ ค่าพารามิเตอร์อำนาจจำแนกข้อสอบและค่าพารามิเตอร์ความสามารถของผู้สอบในโมเดล MIRT อยู่ในรูปของเวคเตอร์มากกว่าสเกลลา และค่าพารามิเตอร์ที่เกี่ยวข้องกับความยากของข้อสอบ (Difficulty-Related Parameter) คือค่าการรวมกันของค่าความยากของข้อสอบและค่าอำนาจจำแนกในแต่ละมิติ ซึ่งค่าอำนาจจำแนกข้อสอบมีความสัมพันธ์ว่ามีความไว (Sensitive) ของความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องเป็นอย่างไร ในขณะที่โมเดล UIRT ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบเป็นสัดส่วนของความชันในโค้ง ICC ณ จุดเปลี่ยนโค้งที่มีความชันสูงสุด (Steepest)

2.4 การหมุนแกนในตำแหน่งที่เหมาะสมและสเกลในตำแหน่งที่เหมาะสม (Rotational Indeterminacy and Scale Indeterminacy)

ในการศึกษาความไม่แปรเปลี่ยนของค่าพารามิเตอร์ตามโมเดล MIRT ประกอบด้วย 2 ส่วนที่จำเป็นต้องให้ความสนใจ นั่นคือ การหมุนแกนในตำแหน่งที่เหมาะสม (Rotational Indeterminacy) และตำแหน่งสเกลที่เหมาะสม (Scale Indeterminacy) โดยมีรายละเอียดดังนี้

2.4.1 การหมุนแกนในตำแหน่งที่เหมาะสม (Rotational Indeterminacy) ในโมเดล MIRT คือ คุณลักษณะ (Characterized) โดยการเชื่อมหรือการเพิ่มการหมุนแกนในตำแหน่งที่เหมาะสม (Rotational Indeterminacy) ของค่าพารามิเตอร์อำนาจจำแนก (a_i) และค่าความสามารถผู้สอบ (θ_j) โดยแกน θ_j สามารถหมุนแกนโดย **Pre-Multiplication** จากเมตริกซ์ T_R , ขณะที่ $a_i^{,R}$ คือ Simultaneous Post-Multiplied โดยเมตริกซ์ T_R^{-1} ดังนั้นความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องของ θ_j ไม่เปลี่ยนแปลง (Hirsch, 1989) ซึ่งสามารถแสดงได้ดังสมการ

$$\theta_j^R = T_R\theta_j \quad \text{.....(2.26)}$$

และ

$$a'^R_i = a_i' T_R^{-1} \quad \text{.....(2.27)}$$

จำนวนคู่ของ θ_j^R และ a_i^R จะมีจำนวนไม่จำกัด โดยให้ $P\left(x_{is}=1 \middle| a_i, d_i, \theta_j\right)$ ไม่เปลี่ยนแปลง การประมาณค่าพารามิเตอร์ตามแนวคิดของ MIRT จะใช้โปรแกรม TESTFACT ที่นำเสนอโดย Wilson et al. ในปี ค.ศ. 1991 เพื่อระบุปัญหา โดยการกำหนด θ ให้มีการกระจายแบบปกติพหุ (Multivariate Normal) มีค่าเฉลี่ยเท่า 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 (MVN(0,1)) ในขั้นตอนสุดท้ายของการประมาณค่า นอกจากนี้ ตามที่มีการประมาณค่า a_i จากหลายๆ สมการ เช่น จากการวิเคราะห์ข้อสอบสารสนเทศเต็มรูปแบบ (Full-Information Item Analysis) จากการศึกษาของ Muraki & Engelhard (1985) รวมถึงจากการศึกษาของ Bock, Gibbons, & Muraki (1988) พบว่า สามารถกำหนดเกณฑ์ในการหมุนแกนได้ เช่น กำหนดการหมุนแกนแบบ Varimax เป็นต้น

2.4.2 ตำแหน่งสเกลที่เหมาะสม (Scale Indeterminacy)

โมเดล MIRT มีลักษณะคล้ายกับโมเดล UIRT นั่นคือ ความไม่แปรเปลี่ยนของการเปลี่ยนจุดกำเนิดและหน่วยความยาว (Shift and Unit Change) ในการประมาณค่าให้มาอยู่ในตำแหน่งสเกลที่เหมาะสม (Scale Indeterminacy) ซึ่งการประมาณค่า θ ในแต่ละมิติจะกำหนดให้อยู่ในรูปคะแนนมาตรฐานโดยมีค่าเฉลี่ยเท่ากับ 0 และ มีความแปรปรวนเท่ากับ 1 หน่วย ภายหลังการประมาณค่าในแต่ละขั้นตอน (Chalson, 1987) สำหรับโครงสร้าง 2 มิติ จุดกำเนิดของแต่ละมิติ θ นั่นคือ $\left(\theta_{j1}, \theta_{j2}\right)$ สามารถที่จะเปลี่ยนตำแหน่งโดยการลบกับค่าเฉลี่ย $\left(\mu_{\theta 1}, \mu_{\theta 2}\right)$ ของการประมาณค่า θ ในแต่ละมิติ และตามที่หน่วยของแต่ละมิติ θ สามารถที่จะทำให้อยู่ในสเกลมาตรฐาน โดยการแปลง

ส่วนเบี่ยงเบนมาตรฐานให้เป็นค่ามาตรฐาน (SD; $\sigma_{\theta 1}, \sigma_{\theta 2}$) ของการประมาณค่า θ (Chalson, 1987; Hirsch, 1989) โดยแสดงได้ดังสมการที่ (2.28) และ (2.29)

$$\theta_{j1}^{\#} = \frac{\theta_{j1} - \mu_{\theta 1}}{\sigma_{\theta 1}} \quad \text{.....(2.28)}$$

และ

$$\theta_{j2}^{\#} = \frac{\theta_{j2} - \mu_{\theta 2}}{\sigma_{\theta 2}} \quad \text{.....(2.29)}$$

ความน่าจะเป็นของการตอบข้อสอบได้ถูกต้องในแต่ละข้อ คือค่าความสามารถของแต่ละบุคคล แทนด้วย θ จะมีค่าไม่เปลี่ยนแปลงถ้าค่าพารามิเตอร์อำนาจจำแนก (a_1, a_2) อยู่บนสเกลมาตรฐานโดยการคูณกับค่าส่วนเบี่ยงเบนมาตรฐาน (SDs) ของ θ ที่ประมาณได้ แสดงได้ดังสมการที่ (2.30), (2.31) และ (2.32)

$$a_{i1}^{\#} = \sigma_{\theta 1} a_{i1} \quad \text{.....(2.30)}$$

และ

$$a_{i2}^{\#} = \sigma_{\theta 2} a_{i2} \quad \text{.....(2.31)}$$

และ d ถูกทำให้อยู่ในสเกลมาตรฐาน (Rescale) โดย

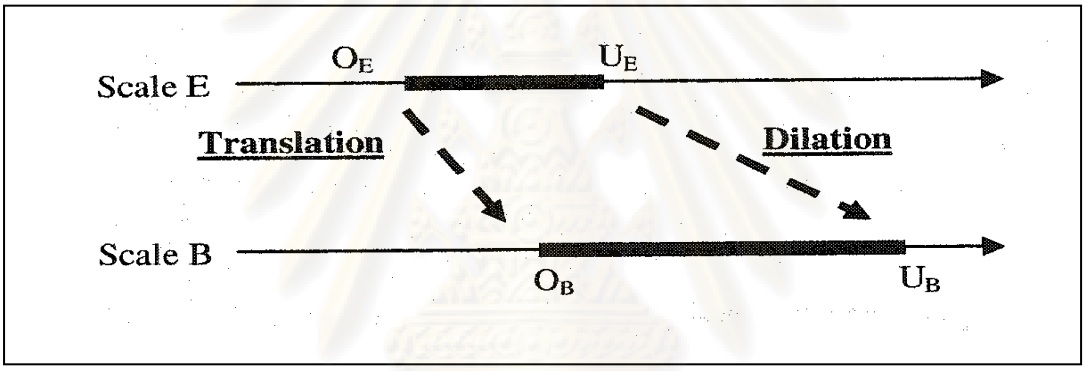
$$d_i^{\#} = d_i + a_{i1}\mu_{\theta 1} + a_{i2}\mu_{\theta 2} \quad \text{.....(2.32)}$$

ตามที่นำเสนอการแปลงคะแนนดังกล่าว จุดกำเนิดของการประมาณค่า θ ในแต่ละมิติจะถูกเปลี่ยนไปยังคะแนนมาตรฐาน (Z Scores) $\left(\theta_{j1}^{\#}, \theta_{j2}^{\#}\right)$ ซึ่งมีค่าเฉลี่ยเท่ากับ 0 และค่าความแปรปรวนเท่ากับ 1 โดยจากสมการที่ (33) และสมการที่ (34) ค่าการแปลงคะแนนของ $a_{i1}^{\#}$ และ $a_{i2}^{\#}$ เท่ากับผลคูณของการประมาณค่า a ที่จุดกำเนิด กับค่าส่วนเบี่ยงเบนมาตรฐาน ($\sigma_{\theta 1}$ หรือ $\sigma_{\theta 2}$) ดังนั้นค่าความแปรปรวนใหม่ที่ได้จากการแปลงค่า a ในแต่ละมิติ มีค่าเท่ากับผลคูณของค่าความแปรปรวนที่จุดกำเนิดกับค่าความแปรปรวนของ θ ที่ได้จากการประมาณค่า

ในปี ค.ศ. 1997 Reckase ได้บ่งชี้ว่า “ถ้าความสัมพันธ์ระหว่างมิติ θ มีค่าเท่ากับ 0 *ดังนั้น ความสัมพันธ์ที่สามารถสังเกตได้ระหว่างคะแนนของข้อสอบจะสามารถนับได้จากค่าพารามิเตอร์ a*” ซึ่งดูคล้ายกับว่าถ้าค่าความแปรปรวนของความสามารถหลายมิติเป็นค่าคงที่โดยมีค่าเท่ากับ 1 และการประมาณค่าพารามิเตอร์ a โดยส่วนใหญ่ได้มาจากโปรแกรมการประมาณค่าตามโมเดล MIRT มีผลให้ค่าความแปรปรวนของจุดกำเนิดแตกต่างกันในแต่ละมิติ θs (Heterogeneous Variances of Multidimensional θs) โดยมีการแจกแจงของ θ แบบปกติพหุ (MVN(0,I)) โดยค่าความแปรปรวนของจุดกำเนิดมีความแตกต่างกันในแต่ละมิติ θs จะยึดจากค่าความแปรปรวน-ค่าความแปรปรวนร่วมของการประมาณค่าพารามิเตอร์อำนาจจำแนกข้อสอบ

จากการนำเสนอสมการความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องดังสมการที่ 2.21 ซึ่งเป็นฟังก์ชันเชิงเส้นตรงของค่าพารามิเตอร์ข้อสอบ (a และ d) และค่าความสามารถ (θ) การแปลง

คะแนนเชิงเส้นตรงของสเกลความสามารถให้อยู่ในสเกลเดียวกัน ซึ่งเป็นสัญลักษณ์สำหรับรูปแบบการเสนอคำตอบ ถ้าค่าพารามิเตอร์ของข้อสอบและค่าพารามิเตอร์ความสามารถถูกแปลงในทิศทางที่สอดคล้องกัน ซึ่งความน่าจะเป็นที่กล่าวถึงคือ ผู้สอบทำข้อสอบได้ถูกต้องนั่นคือการถูกระบุเมื่อสเกลและค่าพารามิเตอร์ของข้อสอบถูกแปลงอย่างถูกต้องและเหมาะสม นั่นเอง คุณสมบัติดังกล่าวเรียกว่าคุณสมบัติของความไม่แปรเปลี่ยน (Kolen & Brennan, 2004; Lord, 1980) ขณะที่การกำหนดตำแหน่งสเกลที่เหมาะสม (Scale Indeterminacy) เป็นตำแหน่งที่ไม่ได้กำหนดจุดกำเนิดและหน่วยของสเกลไว้อย่างเจาะจง (Min, 2003) ขึ้นอยู่กับการค้นพบในตำแหน่งที่เหมาะสม สำหรับแปลงคะแนนในบริบทของการปรับเทียบคะแนนตามโมเดล UIRT (UIRT Equating) จะใช้การหมุนแกนเพื่อใช้ในการปรับเทียบระบบการสรุปอ้างอิง (Comparable Reference System) ซึ่งแสดงได้ดังแผนภาพที่ 2.7 โดยมีลักษณะเช่นเดียวกับการพิจารณาการแปลงคะแนนหรือการปรับเทียบคะแนนในโมเดล MIRT

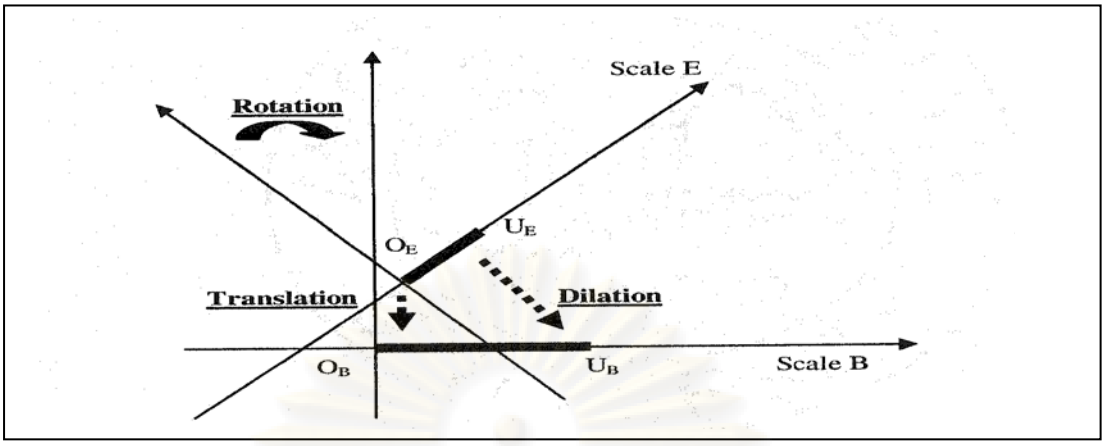


แผนภาพที่ 2.8 ส่วนประกอบของ UIRT Linking

* o คือ ตำแหน่งของจุดกำเนิด, U คือ ความยาวของหน่วย, Subscript E คือ metric ของการแปลงคะแนน และ B คือ จุดมุ่งของ metric (Min, 2003)

จากแผนภาพที่ 2.7 เป็นส่วนประกอบของการเชื่อมโยงคะแนนตามโมเดล UIRT ซึ่งมีลักษณะเช่นเดียวกับการพิจารณาการแปลงคะแนนตามโมเดล MIRT ในบริบทของการปรับเทียบคะแนน จากมุมมองทางด้านเรขาคณิต พบว่าการประมาณค่าเชิงตัวเลขของพารามิเตอร์ข้อสอบตามโมเดล MIRT ขึ้นอยู่กับการกำหนดแกนอ้างอิงหรือ มิติของ θ ซึ่งแสดงได้ดังแผนภาพที่ 2.8

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



แผนภาพที่ 2.9 ส่วนประกอบของ MIRT Linking (2 มิติ) จากการแปลงคะแนนองค์ประกอบ (Composite Transformation): A Rotation, a Translation, and a Central Dilation
(MIRT Linking จากภาพเป็นการศึกษาของ Li ในปี 1997 หน้า 37)

จากแผนภาพที่ 2.8 แสดงถึงเวคเตอร์ขั้นพื้นฐานของการเชื่อมโยงที่ต่างกันจากแบบสอบปรับเทียบคะแนน (Equated Form) ไปยังแบบสอบฐาน (Based Form) ซึ่ง Green (1976) ได้อธิบายแผนภาพที่ 2.8 ไว้ดังนี้

1) แกนของแบบสอบฐานคือ B_1 และ B_2 ส่วนแกนของแบบสอบปรับเทียบคะแนนคือ E_1 และ E_2 สามารถหมุนแกนไปยังแกนให้สอดคล้องกับแบบสอบพื้นฐาน โดยวิธีการ Premultiplying หรือ Postmultiplying โดยเมตริกซ์การหมุนแกนแบบตั้งฉาก (Orthogonal Rotation Matrix) ซึ่งสามารถเขียนแทนด้วยสัญลักษณ์ T (หรือ T^{-1}) ดังได้เคยกล่าวในรายละเอียดของการกำหนดการหมุนแกนที่เหมาะสม (Rotational Indeterminacy)

2) จุดกำเนิดของแบบสอบฐานคือ O_B ส่วนจุดกำเนิดของแบบสอบปรับเทียบคะแนน คือ O_E ซึ่งมีการเปลี่ยนตำแหน่งจาก O_B ไปยัง O_E โดยมีจุดโคออร์ดิเนตของ O_B และ O_E เท่ากับ (0,0) และ (m_1,m_2) ตามลำดับ เมื่อ m_1 คือ ความยาวที่เพิ่มขึ้นจากจุดกำเนิดไปยังมิติที่ 1 นั่นคือจากแบบสอบฐานไปยังแบบสอบปรับเทียบคะแนน ในทำนองเดียวกัน m_2 คือ ความยาวที่เพิ่มขึ้นจากจุดกำเนิดไปยังมิติที่ 2 โดยที่ m_1 และ m_2 คือ สัมประสิทธิ์การแปลงคะแนนที่ใช้ในการแปลงคะแนนจาก O_E ไปยัง O_B

3) หน่วยความยาวของแบบสอบฐานมีค่าเท่ากับส่วนของเส้นตรง O_B ไปยัง U_B ส่วนหน่วยความยาวของแบบสอบปรับเทียบคะแนนมีค่าเท่ากับส่วนของเส้นตรง O_E ไปยัง U_E ซึ่งสัดส่วนของหน่วยความยาวของแบบสอบปรับเทียบคะแนนต่อหน่วยความยาวของแบบสอบฐานเป็นค่าสัมประสิทธิ์การเลื่อนแกน (Dilation Coefficient) แทนด้วยสัญลักษณ์ k หรือเรียกว่าค่าพารามิเตอร์การเลื่อนแกน

การแปลงคะแนนองค์ประกอบ (Composite Transformation) ของสเกล และการหมุนแกนการแปลงคะแนน (Rotation Transformation) สามารถทำให้ชัดเจนขึ้น โดยการใช้เมตริกซ์การหมุนแกน (Metrix Rotation) นั่นคือ ถ้าโมเดลที่เลือกสอดคล้องกับข้อมูลเชิงประจักษ์ $a_{iE}'\theta_{jE} + d_{iE}$ (ดังแสดงในสมการที่ (2.33)) ยังมีค่าไม่เปลี่ยนแปลง ทั้งค่าพารามิเตอร์ของข้อสอบ $\left(a_{iE}', d_{iE}\right)$และค่าพารามิเตอร์ของผู้สอบ $\left(\theta_{jE}\right)$ (McKinley & Mills, 1985; Tam & Li, 1997) แสดงสมการในการแปลงคะแนนดังนี้

$$a^{\#}_{iB} = k a^{'}_{iE} T, \quad \text{.....(2.33)}$$

$$d^{*}_{iB} = d_{iE} + \left(a^{'}_{iE} T\right) m, \quad \text{.....(2.34)}$$

และ

$$\theta^{*}_{iB} = (1/k)\left(T^{-1}\theta_{iE} - m\right) \quad \text{.....(2.35)}$$

เมื่อ * แทน ค่าของการแปลงคะแนนจาก Equated Scale ไปยัง Base Scale

k แทน ค่าพารามิเตอร์ Dilation

T แทน เมตริกซ์ของการหมุนแกนแบบออร์โธโกนอล (Ortogonal Rotation Matrix)

M แทน เวคเตอร์การแปลงคะแนน (Translation Vector)

อย่างไรก็ตาม จากการศึกษาของ Oshima et al. (1997) มีส่วนที่ขัดแย้งในเรื่องของสมการการแปลงคะแนน สามารถแสดงได้ดังสมการ

$$a^{\#}_{iB} = \left(A^{-1}\right) a_{iE} \quad \text{.....(2.36)}$$

$$d^{*}_{iB} = d_{iE} - \left(a^{'}_{iE} A^{-1}\right)\beta \quad \text{.....(2.37)}$$

และ

$$\theta^{*}_{iB} = A\theta_{jE} - \beta \quad \text{.....(2.38)}$$

จากสมการดังกล่าวเห็นได้ว่า เมตริกซ์การหมุนแกน A ซึ่งปรับให้เป็นค่าความแปรปรวนและค่าความแปรปรวนร่วมของมิติ θ และเวคเตอร์การแปลงคะแนน β จะการเลื่อนโครงสร้าง (Re-Shifts) จากจุดกำเนิดในแต่ละมิติ θ จากสมการบ่งชี้ว่าไม่มีค่าพารามิเตอร์การเลื่อนแกน (Dilation; k) ในสมการปรับเทียบตามโมเดล MIRT ของ Oshima et al. ถ้าการหมุนแกนของเมตริกซ์เป็นแบบไม่ตั้งฉากจะส่งผลให้ตำแหน่งเชิงสัมพัทธ์ของค่าพารามิเตอร์อำนาจจำแนกทั้งสองค่าของแบบสอบปรับเทียบคะแนนผิดรูปไปจากความเป็นจริงได้ในขณะที่แปลงคะแนนไปสู่แบบสอบฐาน

2.5 การประมาณค่าการแปลงค่าพารามิเตอร์ตามโมเดล MIRT (Estimating MIRT Transformation Parameters)

เทคนิคการประมาณค่าการแปลงตามโมเดล MIRT จะมีการประมาณค่าใน 2 ส่วนหลัก คือ 1) การประมาณค่าเมตริกซ์การหมุนแกน (Rotation Matrix) ใช้สำหรับประมาณค่าอำนาจจำแนกข้อสอบในแบบสอบปรับเทียบคะแนน และ 2) การประมาณค่าสัมประสิทธิ์การสร้างสเกล (Scaling Coefficient) มีตัวที่ใช้ในการประมาณค่า 4 ชนิด คือ Matching Test Characteristic Surfaces (MTCS), Least Squares เพื่อใช้ในประมาณการแปลงค่าพารามิเตอร์ และ Ratio of Eigenvalue หรือ Ratio of Trace สำหรับใช้ในการประมาณค่าพารามิเตอร์การเลื่อนแกน (Dilation) โดยมีรายละเอียดดังนี้

2.5.1 หลักการของการประมาณค่าเมตริกซ์การหมุนแกน (Estimating the Rotation Matrix)

Ordinary Orthogonal Procrustes Rotation (SchÖnemann, 1966) ใช้สำหรับการหมุนแกนในการประมาณค่าอำนาจจำแนกข้อสอบในแบบสอบปรับเทียบคะแนน นั่นคือเป็นผลรวมของความแตกต่างยกกำลังสองระหว่างค่าอำนาจจำแนกข้อสอบที่ได้จากการประมาณค่าในแบบสอบ

ปรับเทียบคะแนนกับแบบสอบฐาน ในแต่ละคู่ (ระหว่าง a_{iE} กับ a_{iB}) โดยต้องมีค่าความแตกต่างน้อยที่สุดนั่นเอง สามารถแสดงได้ดังสมการที่ (2.39)

$$E_1 = A_E T - A_B \quad \text{.....(2.39)}$$

เมื่อ E_1 แทน เมตริกซ์ส่วนที่เหลือ (Residual Matrix)

T แทน เมตริกซ์การแปลงคะแนน (Transformation Matrix)

A_E และ A_B แทน เมตริกซ์ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบที่มีขนาด L x m สำหรับกลุ่มปรับเทียบคะแนนและสำหรับกลุ่มพื้นฐาน

โดยที่ L แทน จำนวนของข้อสอบร่วม

2.5.2 หลักการพื้นฐานในการประมาณค่าสัมประสิทธิ์ของสเกล (Estimating the Scaling Coefficients; m_1, m_2, k)

วิธีการที่ใช้ในการประมาณค่าสัมประสิทธิ์ m_1, m_2, และ k โดยทั่วไปนิยมใช้ 2 วิธี วิธีการแรกเป็นการเชื่อมกับพื้นผิวการตอบแบบสอบ (Matching the Test Response Surface) ส่วนวิธีการเป็นวิธีกำลังสองน้อยสุด (Least Square) เพื่อใช้สำหรับการแปลงค่าพารามิเตอร์ในการเชื่อมกับสัดส่วนของค่าไอเกน (Eigenvalues) หรือสัดส่วนของจุดที่ตัดกัน (Trace) ซึ่งเป็นค่าพารามิเตอร์การเลื่อนแกน (Dilation) โดยมีรายละเอียดในแต่ละเทคนิควิธี ดังนี้

(1) การเชื่อมพื้นผิวการตอบสนองของแบบสอบ (Matching the Test Response Surface; MTCS)

ฟังก์ชันการตอบสนองแบบสอบของโมเดล MIRT คือ พื้นผิวการตอบแบบสอบที่ (Test Response Surface; TRSs) เป็นวิธีการของการรวมความน่าจะเป็นของการตอบข้อสอบได้ถูกต้องของข้อสอบร่วม มีค่าเท่ากับจำนวนที่คาดว่าจะตอบข้อสอบได้ถูกต้อง (Expected Number Correct; ENC) ซึ่งวิธีการ Matching TRSs (MTRs) เป็นแนวคิดที่แผ่ขยายมาจากแนวคิดของ Stocking & Lord's (1983) เป็นกระบวนการที่ Oshima et al. (1997) นำมาใช้ปรับเทียบคะแนนตามโมเดล MIRT ในกรณีที่มี 2 มิติ การแปลงคะแนนเชิงเส้นของการตอบข้อสอบได้ถูกต้องของสเกลจากข้อสอบร่วมซึ่งแทรกอยู่ในแบบสอบที่แตกต่างกัน 2 ฉบับ เพื่อมีจุดมุ่งหมายที่จะสร้าง ENC ให้อยู่ในสเกลเดียวกัน สำหรับผู้สอบคนที่ j ถ้าทราบค่าสัมประสิทธิ์ของ m_1, m_2, และ k ในทางปฏิบัติ ต้องการให้ค่าเฉลี่ยยกกำลังสองของความแตกต่างระหว่าง TRSs มีค่าน้อยที่สุดเท่าที่จะเป็นไปได้ โดยสามารถแสดงฟังก์ชันได้ดังสมการที่ (2.40)

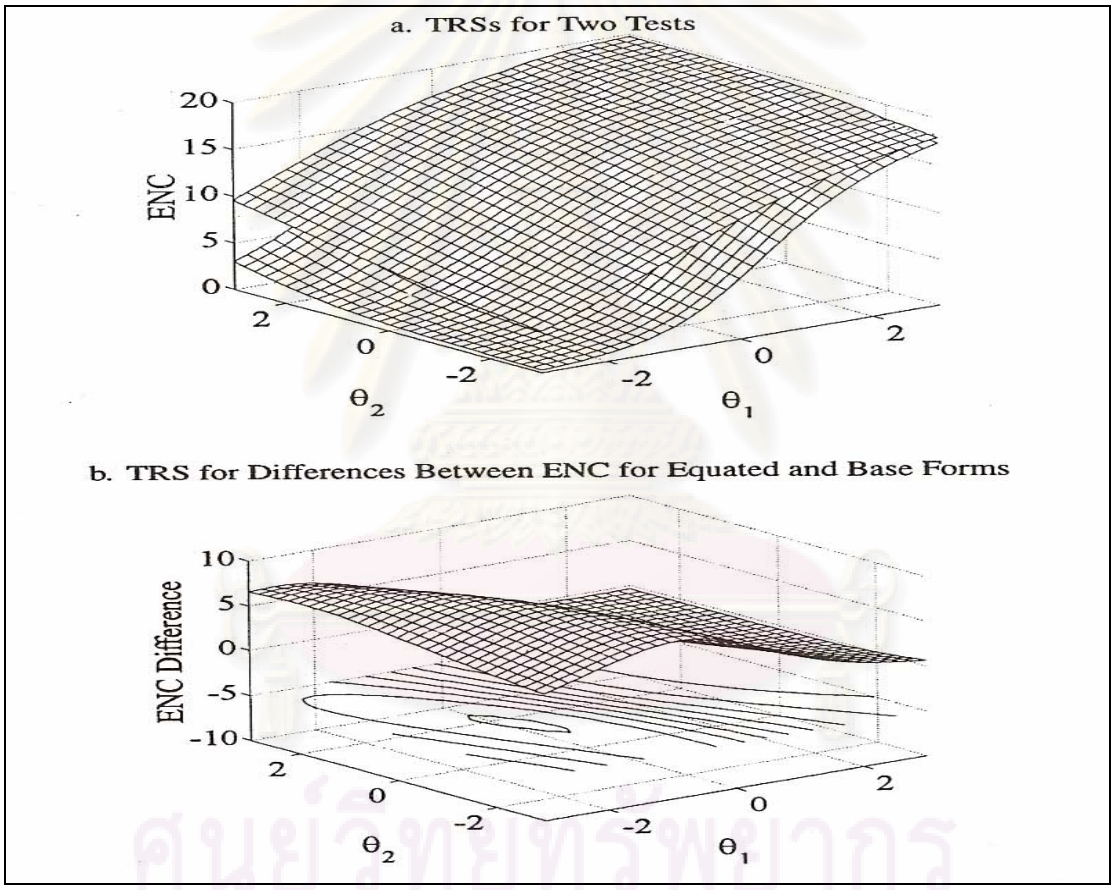
$$F = \frac{1}{H}\left[t\left(\theta_{j1},\theta_{j2}\right) - t^{*}\left(\theta_{j1},\theta_{j2}\right)\right]^2 \quad \text{.....(2.40)}$$

เมื่อ $t\left(\theta_{j1},\theta_{j2}\right)$ และ $t^{*}\left(\theta_{j1},\theta_{j2}\right)$ แทน ค่า Expected Number Correct หรือ ENC ของผู้สอบบนเซตของข้อสอบร่วมของแบบสอบพื้นฐานและแบบสอบปรับเทียบคะแนนตามลำดับ

H แทน จำนวนของ Grid Points

สำหรับกรณีที่มีการศึกษา 2 มิติ H มีค่าเท่ากับ h^2 ถ้า θ ในแต่ละมิติถูกแปลงไปยัง h points โดยส่วนใหญ่กำหนดสเกลของ θ อยู่ในช่วง -4 ถึง 4 เช่น -4, -3.8, -3.6,…,3.6, 3.8, 4 เป็นต้น ซึ่งการรวมองค์ประกอบ H ในมิติ θ ทั้งหมด สามารถคำนวณได้จากสมการที่ (2.40) ซึ่งการแปลงคะแนนของการประมาณค่าพารามิเตอร์จะมีความสัมพันธ์อย่างคงที่ (Relatively Stable) เมื่อ H มีขนาดมากกว่า 100

ส่วนการประมาณค่า m_1, m_2 และ k ให้ F มีค่าน้อยๆ สามารถทำได้โดยการแปลงสมการที่ (2.41) ซึ่งมีความเกี่ยวข้องกับ m_1, m_2 และ k ซึ่งกำหนดให้สมการบางส่วนมีค่าเท่ากับ 0 และใช้กระบวนการของ Newton-Raphson (Hambleton & Swaminathan, 1985) คำนวณค่า m_1, m_2 และ k



แผนภาพที่ 2.10 Test Response Surfaces (TRs) (Ackerman, Gierl & Walker, 2003)

จากแผนภาพที่ 2.10a เป็นตัวอย่างของกระบวนการดังที่กล่าวมาแล้วข้างต้น จะเห็นได้ว่า Lower TRs อยู่บนพื้นฐานของการประมาณค่าพารามิเตอร์ 20 ข้อแรก จากแบบสอบ ACT Form 24B (Reckase, 1985) ส่วน Upper TRs เป็นกราฟที่ได้จากการใช้แบบสอบชุดเดียวกับ Lower TRs แต่การประมาณค่าพารามิเตอร์ของข้อสอบจะถูกปรับให้อยู่ในค่ามาตรฐาน (Calibrated) จากผู้สอบอีกกลุ่มหนึ่ง (Rotation Indeterminacy of Those Estimates d Parameters was Assumed to be Resolved) จะมีการพล็อตค่า Upper TRs โดยใช้การประมาณ

ค่าพารามิเตอร์ d โดยการบวกด้วย 1 ทุกค่าของ d ส่วนการประมาณค่าพารามิเตอร์ a คูณด้วย 0.5 ทุกค่าของ a โดยคาดหวังว่าการกำหนดค่าพารามิเตอร์ในสเกล (m_1=-1, m_2=-1 และ k=2) นั้น ต้องทำให้ค่าความแตกต่างยกกำลังสองโดยเฉลี่ย (Average Squared Difference)ระหว่าง TRss มีค่าน้อยที่สุด

สำหรับแผนภาพที่ 2.10b แสดงถึงพื้นผิวของความแตกต่างระหว่าง ENCs ของ Equated Form กับ Based Form ซึ่งเรียกว่า “Contour Plot” ซึ่งจากแผนภาพได้แสดงให้เห็นถึงทิศทางและตำแหน่งของความแตกต่างก่อนและหลังการเชื่อมโยงคะแนน (Linking) ของแบบสอบทั้งสองฉบับ

(2) วิธีกำลังสองน้อยที่สุดสำหรับการประมาณค่าพารามิเตอร์การแปลงคะแนน (Least Square for Estimating Transformation Parameters) และวิธีการศึกษาสัดส่วนของค่าไอเกนสำหรับการประมาณค่าพารามิเตอร์ Dilation (Ratio of Eigenvalues for Estimating the Dilation Parameter)

ค่าของ m_1, m_2 และ k สามารถประมาณค่าแยกส่วนกันได้ ซึ่งวิธีการนี้เรียกว่า “**Least Square**” ใช้สำหรับค้นหาตัวประมาณค่า (Estimators) ของ m_1 และ m_2 ในแต่ละตัวอย่างที่สามารถสังเกตได้ (d_{iE}, d_{iB}) โดยค่าเบี่ยงเบนของ d_{iB} จะพิจารณาจากค่าการแปลงคะแนนของ d_{iE} (d^{*}_{iE}) สามารถแสดงได้ดังสมการที่ (2.37) โดยเกณฑ์ที่ใช้สามารถพิจารณาได้จากค่า Q ดังสมการที่ (2.41)

$$Q = \sum_{i=1}^{L} \left(d_{iB} - d^{*}_{iB} \right)^2 \quad \text{.....(2.41)}$$

ค่าของ m_1 และ m_2 ที่มีค่า Q น้อย พิจารณาจากความแตกต่างในสมการที่ (2.41) เมื่อกำหนดให้สมการบางส่วนมีค่าเท่ากับ 0 ค่าของ m_1 และ m_2 สามารถประมาณค่าได้จาก 2 สมการดังที่ได้กล่าวมาแล้ว

ค่าพารามิเตอร์การเลื่อนแกน (Dilation; k) สามารถประมาณค่าได้จากสัดส่วนของเทคนิคค่าไอเกน (Eigenvalues) ซึ่งเป็นสัดส่วนของรากที่สองของค่าไอเกนสูงสุดของ $B'B$ กับสัดส่วนของรากที่สองของค่าไอเกนสูงสุดของ E’E ซึ่งค่ารากที่สองที่ไม่เป็นลบของค่าไอเกนของ $B'B$ เรียกว่า “Singular Values” ของ B ($B=A'_BA_B$; แทนด้วยสัญลักษณ์ Sig(B)) เช่นเดียวกับค่ารากที่สองที่ไม่เป็นลบของค่าไอเกนของ $E'E$ เรียกว่า “Singular Values” ของ E ($E=A'_EA_E$; แทนด้วยสัญลักษณ์ Sig(E) สามารถแสดงสมการของ k ได้ดังนี้

$$k = \frac{Max\left[Sig(B)\right]}{Max\left[Sig(E)\right]} \quad \text{.....(2.42)}$$

เมื่อ Max แทน ค่าของฟังก์ชันสูงสุด

สารสนเทศที่ได้เกี่ยวกับ B มาจากคุณสมบัติสำคัญของ B- λ_B I เมื่อ λ_B แทน Characteristics Roots หรือค่าไอเกนของ B ซึ่งค่าสูงสุดของ λ_B สามารถอธิบายถึงความแปรปรวนมากที่สุดของ B (Tatsuoka & Lohnes, 1988) ส่วนค่ารากที่สองของค่าไอเกนที่มากที่สุดของ B เป็นการบ่งบอกถึงดัชนีความเที่ยงในสเกล ($Max[Sig(B)]$) ในทำนองเดียวกันสารสนเทศที่ได้เกี่ยวกับ E ทำให้ทราบถึงดัชนีความเที่ยงของสเกลด้วยเช่นกัน ($Max[Sig(E)]$) ซึ่งเป็นผลให้อัตราส่วนของ $Max[Sig(B)]$ ต่อ $Max[Sig(E)]$ เป็นค่าที่ดีในการประมาณค่าพารามิเตอร์การเลื่อนแกน (Dilation) ได้

(3) การประมาณค่าพารามิเตอร์การเลื่อนแกนด้วยวิธีกำลังสองน้อยที่สุด และสัดส่วนของจุดที่ตัดกัน (Least Square Estimation for Dilation Parameter & Ratio of Trace for the Dilation Parameter Estimation)

ในปี ค.ศ. 1970 SchÖnemann & Carroll ได้เสนอเทคนิควิธี Least Square ในการประมาณค่า k ซึ่งพัฒนามาจากความสอดคล้องของเมตริกซ์หนึ่งไปยังเมตริกซ์อื่นๆ ภายใต้เงื่อนไข 3 ประการ คือ 1) Rotation Matrix, 2) Translation Vector หรือ 3) Central Dilation Vector โดยผลรวมของความคลาดเคลื่อนยกกำลังสองของเมตริกซ์เศษที่เหลือ E_2 ที่น้อยที่สุดสามารถแสดงได้ดังสมการที่ (2.42)

$$E_2 = \left(kA_E T\right) - A_B \quad \text{.....(2.42)}$$

เมื่อ k แทนสัมประสิทธิ์การเปลี่ยนแปลงของหน่วยความยาว (Unit Change Coefficient)

เมื่อ k แทน สัมประสิทธิ์การเปลี่ยนแปลงหน่วยความยาว และ Translation Transformation ถูกตัดจากสูตรเดิม เนื่องจากค่าพารามิเตอร์อำนาจจำแนกตามโมเดล MIRT ไม่สามารถแสดงให้เห็นถึงสารสนเทศใดๆเกี่ยวกับ Translation Vector หรือ Original Shift Coefficients ได้ ซึ่งเทคนิค Least Square โดยทั่วไปจะใช้สำหรับ **Nonmetric Multidimensional Scaling** แต่สามารถนำมาประยุกต์ใช้กับการเชื่อมโยงคะแนนตาม MIRT ได้ โดยสมการประมาณค่า T และ k สามารถประมาณได้ในคราวเดียวกัน จากการประมาณค่า T ในสมการที่ (2.42) ซึ่งเป็นอิสระจากการประมาณค่า k (SchÖnemann & Carroll, 1970) สามารถสรุปขั้นตอนในการประมาณค่า T และ k ได้ทั้งหมด 4 ขั้นตอน คือ 1) ศูนย์กลางของส่วนประกอบทั้งหมด (Center All Element) ใน A_E จะสร้างให้เป็ ”เมตริกซ์ศูนย์กลาง” แทนด้วยสัญลักษณ์ A_{CE} 2) ศูนย์กลางของส่วนประกอบทั้งหมด (Center All Element) ใน A_B จะสร้างให้เป็น ”เมตริกซ์ศูนย์กลาง” แทนด้วยสัญลักษณ์ A_{CB} 3) การหมุนแกนมาตรฐานแบบตั้งฉาก (Standard Orthogonal Procrusters Rotation) จะให้ค่า T และ 4) คำนวณค่า k ซึ่งสามารถแสดงได้ดังสมการที่ (2.43)

$$k = \frac{trace\left(T' A'_{CE}\right) A_{CB}}{trace\left(A'_{CE}\right) A_{CE}} \quad \text{.....(2.43)}$$

จะเห็นว่าการประมาณค่าพารามิเตอร์ทั้ง 3 ส่วน คือ m_1, m_2 และ k สามารถใช้เทคนิคของ MIRT ในการเชื่อมโยงคะแนนเพื่อใช้ในการแปลงค่าจากกลุ่มปรับเทียบคะแนน ไปยังกลุ่มฐาน ซึ่งการแปลงค่าพารามิเตอร์สามารถใช้โปรแกรม MDEQUATE (Li, 1996) ซึ่งเขียนเป็นภาษา MATLAB เวอร์ชัน 5.1 (The Math Works, Inc., 1997)

2.6 การกระจายความคลาดเคลื่อนสำหรับการประมาณค่าพารามิเตอร์ (Error Distribution for Item Parameter Estimates)

ขนาดของกลุ่มตัวอย่างซึ่งส่งผลต่อรูปร่างของการกระจายค่าความสามารถ (θ) และการประมาณค่าพารามิเตอร์ของข้อสอบ เป็นสาเหตุต่อความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ (Hambleton, Jones, & Roger, 1993; Stocking, 1990; Thissen & Wainer, 1982) โดย Thissen & Wainer (1982) ได้พัฒนาสูตรทางด้านคณิตศาสตร์เพื่อแสดงความสัมพันธ์ของโมเดลที่มีการตรวจให้คะแนนแบบ 2 ระหว่าง UIRT กับ MIRT เพื่อพิจารณาถึงความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์

สำหรับข้อสอบข้อที่ i ความเป็นไปได้ (Likelihood) ของการตอบข้อสอบสามารถสังเกตได้สำหรับผู้สอบจำนวน N คนที่เป็นอิสระต่อกัน ดังสมการที่ (2.44)

$$L=\prod_{j=1}^{N} P_j^{U}\left(1-P_j\right)^{1-U} \qquad \text{.....(2.44)}$$

เมื่อ P แทน ค่าที่คำนวณได้จาก M2PLM

u แทนผลการตอบข้ออบ ตอบผิด = 0, ตอบถูก = 1

เมื่อมีการปรับสเกลฟังก์ชันความน่าจะเป็นให้สามารถพิจารณาได้ง่ายขึ้น โดยใช้ลอกาลิทึมเรียกว่า ฟังก์ชัน log ของความน่าจะเป็น (Loglikelihood) สามารถแสดงได้ดังสมการที่ (2.45)

$$\log L=\sum\left[u\log\left(P_j\right)+\left(1-u\right)\log\left(1-P_j\right)\right] \qquad \text{.....(2.45)}$$

การประมาณค่าที่เป็นไปได้สูงสุดของพารามิเตอร์แต่ละค่า (a_i, d_i) โดยกำหนดให้ Partial Derivatives จากสมการที่ (2.45) มีค่าเท่ากับ 0

เมื่อกำหนดให้ ξ แทนค่าพารามิเตอร์ของข้อสอบแบบพหุมิติชนิด 2 พารามิเตอร์ (M2PLM) นั่นคือ a_i และ d_i โดย θ มีการกระจายแบบปกติพหุ (MVN(0,1)) สำหรับแต่ละ ξ_s และ ξ_t จะได้ ค่าที่คาดหวังเป็นลบ (Negative Expected Value) ของการแปลงลำดับที่สอง (Second Derivative) ในฟังก์ชัน log ของความน่าจะเป็น (Loglikelihood) (จากสมการที่ (2.45)) สามารถแสดงได้ดังสมการที่ (2.46)

$$-E\left(\frac{\partial^2 LogL}{\partial\xi_s\xi_t}\right)=N\int_{-\alpha}^{\alpha}\int_{-\alpha}^{\alpha}\left\{\left(\frac{1}{PQ}\right)\left[\frac{\partial P(\theta)}{\partial\xi_s}\frac{\partial P(\theta)}{\partial\xi_t}\right]\right\}\phi_i(\theta)d\theta_1 d\theta_2 \qquad \text{.....(2.46)}$$

เมื่อ E แทน Expectation และ Q = P-1 (Thissen & Wainer, 1982)

จากสมการที่ (2.46) ต้องมีการแปลง P(θ) ให้สัมพันธ์กับค่าพารามิเตอร์ ซึ่งสามารถสับเปลี่ยน ในรูปแบบของเมตริกซ์สารสนเทศที่มีขนาด 3x3 (สำหรับ M2PLM) เพื่อให้สอดคล้องกับค่าพารามิเตอร์ d, a_1 และ a_2 ส่วนอินเวอร์สของเมตริกซ์สารสนเทศ คือ เมตริกซ์ความแปรปรวน-ความแปรปรวนร่วม Asymtotic ของค่าพารามิเตอร์ทั้ง 3 ค่า ส่วนค่ารากที่สองของ Diagonal Elements ของเมตริกซ์ความแปรปรวน-ความแปรปรวนร่วม คือ ค่าความคลาดเคลื่อนมาตรฐาน Asymptotic ของค่าพารามิเตอร์

การประมาณค่าของตัวเลขในสมการที่ (2.46) สามารถที่จะคำนวณได้โดยวิธีการหาพื้นที่สี่เหลี่ยมจัตุรัส (Two-Dimensional Gauss-Hermite Quadrature) ในกรณีที่ศึกษา 2 มิติ สามารถแสดงได้ดังสมการที่ (2.47)

$$N\sum\sum\left\{\left(\frac{1}{PQ}\right)\left[\frac{\partial P(x)}{\partial \xi_s}\frac{\partial P(x)}{\partial \xi_t}\right]\right\}A\left(X_{q1}\right)A\left(X_{q2}\right) \quad \text{.....(2.47)}$$

เมื่อ X แทน Quadrature Point 1 ใน 2 ของมิติ θ

Q แทน จำนวนของ Quadrature ในมิติที่ศึกษา

A(X) แทน ค่าน้ำหนักความสอดคล้องของ Quadrature

โดยทั่วไปนิยมใช้ 10 Quadrature Point ในแต่ละมิติ

กล่าวโดยสรุป การเชื่อมโยงคะแนนตามโมเดล MIRT โดยทั่วไปนิยมประมาณค่าพารามิเตอร์โดยใช้โมเดลโลจิสติกแบบพหุมิติ คล้ายกับโมเดลโลจิสติกตามโมเดล UIRT ในที่นี้ผู้วิจัยได้นำเสนอเฉพาะเทคนิคการเชื่อมโยงคะแนนสำหรับข้อมูลที่มีการตรวจคะแนนแบบ 2 ค่า ซึ่งในการศึกษาความไม่แปรเปลี่ยนของค่าพารามิเตอร์ ตามโมเดล MIRT จะมีอยู่ 2 ส่วนหลักที่จำเป็นต้องให้ความสนใจ นั่นคือการกำหนดการหมุนแกนที่เหมาะสม (Rotational Indeterminacy) และการกำหนดตำแหน่งสเกลที่เหมาะสม (Scale Indeterminacy) โดยนำเสนอวิธีการเชื่อมโยงคะแนนตามวิธีของเวคเตอร์ (Vector Approach) ซึ่งเป็นการศึกษามุมมองทางด้านเรขาคณิต พบว่าการประมาณค่าเชิงตัวเลขของค่าพารามิเตอร์ข้อสอบตามโมเดล MIRT ขึ้นอยู่กับการกำหนดระบบแกนอ้างอิง (Reference System) หรือมิติของความสามารถ (θ)

เมื่อพิจารณาเทคนิคของการประมาณค่าการแปลงค่าพารามิเตอร์ตามโมเดล MIRT จะมีการประมาณค่าใน 2 ส่วนหลัก คือ 1) การประมาณค่าการหมุนแกน (Rotation Matrix) ใช้สำหรับหมุนแกนเพื่อประมาณค่าอำนาจจำแนกข้อสอบในแบบสอบปรับเทียบคะแนน และ 2) การประมาณค่าสัมประสิทธิ์การสร้างสเกล (Scaling Coefficient; m_1, m_2 & k) ซึ่งวิธีการที่ใช้ในการประมาณค่าสัมประสิทธิ์ m_1, m_2, และ k โดยทั่วไปนิยมใช้ 2 วิธีแรกเป็นการเชื่อมกับพื้นผิวการตอบแบบสอบ (Matching the Test Response Surface) ส่วนวิธีที่สองเป็นการประมาณค่ากำลังสองน้อยสุด (Least Square) ใช้สำหรับการแปลงค่าพารามิเตอร์เพื่อเชื่อมกับสัดส่วนของค่าไอเกน (Eigenvalues) หรือสัดส่วนของจุดที่ตัดกัน (Trace) ซึ่งเป็นค่าพารามิเตอร์การเลื่อนแกน (Dilation) โดยการศึกษาการกระจายความคลาดเคลื่อนสำหรับประมาณค่าพารามิเตอร์ พบว่า ขนาดกลุ่มตัวอย่างมีผลต่อรูปร่างของการกระจาย θ และคุณลักษณะของข้อสอบเป็นสาเหตุสำคัญที่ส่งผลต่อความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ (Hambleton, Jones, & Roger, 1993; Stocking,

1990; Thissen & Wainer, 1982) โดย Thissen & Wainer (1982) ได้พัฒนาสูตรทางด้านคณิตศาสตร์เพื่อแสดงความสัมพันธ์ระหว่างโมเดล UIRT ซึ่งสามารถนำมาประยุกต์ใช้สำหรับโมเดล MIRT ที่มีการตรวจให้คะแนนแบบ 2 ค่า เพื่อพิจารณาถึงความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์

หัวข้อต่อไป เป็นหัวใจสำคัญของการศึกษาการเชื่อมโยงคะแนนสำหรับการวิจัยครั้งนี้คือ การกล่าวถึง วิธีการเชื่อมโยงคะแนนตามโมเดล MIRT (MIRT Linking Method) และโครงสร้างมิติความสามารถ (Dimensional Structure) โดยมีรายละเอียดดังต่อไปนี้

2.7 วิธีการเชื่อมโยงคะแนนตามโมเดล MIRT (MIRT Linking Method)

กฎของการเชื่อมโยงคะแนนเพื่อใช้ประมาณค่าพารามิเตอร์ของข้อสอบจากการประมาณค่าที่แยกกันแต่มีค่าความสามารถร่วมกัน โดยตั้งอยู่บนข้อตกลงเบื้องต้นของความไม่เท่าเทียมกันของ 2 กลุ่มผู้สอบแต่มีข้อสอบร่วมกัน จุดมุ่งหมายเพื่อที่จะสร้างการประมาณค่าพารามิเตอร์ 2 ชุด ที่มีลักษณะ “เหมือนกัน (Similar)” เพื่อแปลงค่าพารามิเตอร์จากกลุ่มหนึ่งไปยังอีกกลุ่มหนึ่ง ซึ่งเกี่ยวข้องกับความไม่แปรเปลี่ยนของสเกล (Scale Invarience) เป็นคุณสมบัติของ IRT ในการแปลงเชิงเส้นตรงในผลของสเกลความสามารถให้อยู่ในค่าเดียวกัน (Same Exponent Value) (ดังที่แสดงในสมการที่ 2.21) เพื่อให้รูปแบบการตอบข้อสอบ ถ้าค่าพารามิเตอร์ข้อสอบและค่าพารามิเตอร์ความสามารถถูกแปลงไปยังทิศทางที่สอดคลองกัน นั่นหมายความว่า ความน่าจะเป็นที่ผู้สอบจะทำข้อสอบข้อนั้นได้ถูกต้องเมื่อสเกลและค่าพารามิเตอร์ของข้อสอบได้ถูกแปลงคะแนนได้อย่างเหมาะสม

ในทางปฏิบัติแล้วการประมาณค่าพารามิเตอร์ข้อสอบสามารถทำให้คล้ายกันได้ แต่ไม่สามารถระบุได้ เนื่องจากความแตกต่างของทั้ง 2 เซต ไม่เพียงแต่การกำหนดโมเดลที่เหมาะสมเท่านั้นแต่ขึ้นอยู่กับชนิดที่ใช้ในการเชื่อมโยงคะแนนที่ใช้เป็นเงื่อนไขในการดำเนินการแปลงคะแนน

จากที่ได้กล่าวไว้แล้วว่า Li & Lissitz (2000) ได้ระบุการกำหนดตำแหน่งที่เหมาะสม (Indeterminacy) ไว้ 3 ลักษณะเพื่อใช้เป็นวิธีการในการปรับเทียบคะแนน (MIRT Equting) เพื่อให้เกิดความถูกต้อง นั่นคือ 1) การกำหนดการหมุนแกนที่เหมาะสม (Rotational Indeterminacy) 2) การกำหนดหน่วยที่เหมาะสม (Unit Indeterminacy) และ 3) การกำหนดจุดกำเนิดที่เหมาะสม (Origin Indeterminacy) โดยการหมุนแกนที่เหมาะสมมีวิธีในการหมุนแกนของแบบสอบปรับเทียบคะแนนเพื่อให้เชื่อมต่อกับแบบสอบฐาน สำหรับการกำหนดขนาดของหน่วยที่เหมาะสม (Unit Indeterminacy) เพื่อใช้ในการระบุแบบสอบปรับเทียบคะแนนให้มีการหดหรือขยายเพื่อให้เชื่อมกับหน่วยของแบบสอบฐาน ส่วนการกำหนดจุดกำเนิดที่เหมาะสม (Origin Indeterminacy) สามารถกระทำได้โดยการแปลงจุดกำเนิดที่มีการระบุมิติที่หลากหลายของแบบสอบปรับเทียบคะแนนไปยังจุดกำเนิดของแบบสอบฐาน

วิธีการในการเชื่อมโยงคะแนนมีหลากหลายวิธี (Hirsch, 1989; Li & Lissitz, 2000; Thompson, Nering, & Davy, 1997; Oshima, Davy, & Lee, 2000; Min, 2003; Reckase & Martinue (2004)) ซึ่งมีการพัฒนามาตามลำดับตั้งแต่อดีตถึงปัจจุบัน ซึ่งแต่ละวิธีนิยมออกแบบโดยใช้ข้อสอบร่วมเพื่อเป็นเครื่องมือในการเชื่อมโยงแบบสอบที่มีหลายฉบับให้สามารถอยู่บนสเกลที่สามารถเปรียบเทียบกันได้ อย่างไรก็ตามในแต่ละวิธีของการเชื่อมโยงคะแนนจะมีวิธีการที่เป็นเอกลักษณ์เฉพาะของ

คุณลักษณะของค่าสถิติ (Statistical Characteristics) และมีเกณฑ์ที่มีความเหมาะสมมากที่สุด (Optimization Criteria) ในแต่ละวิธี เช่น อะไรที่มีผลทำให้วิธีการกำหนดตำแหน่งที่เหมาะสมทั้ง 3 ลักษณะมีค่ามากที่สุดหรือน้อยที่สุด เป็นต้น โดยมีรายละเอียดของแต่ละวิธี ดังต่อไปนี้

1.วิธีการของ Hirsch (Hirsch's Method)

Hirsch (1988,1989) ได้มีการพัฒนาวิธีการเชื่อมโยงคะแนนตามโมเดล MIRT บนพื้นฐานของการออกแบบโดยใช้ผู้สอบร่วม (Common Examinee Design) ประกอบด้วย 3 ขั้นตอนของการแปลงคะแนน คือ

ขั้นตอนที่ 1. การค้นหาเวคเตอร์ตั้งฉากที่เป็นแกนหลักหรือส่วนสำคัญ (Orthogonal Basis Vectors)

ขั้นตอนที่ 2. เมตริกซ์การหมุนแกนแบบตั้งฉาก (Orthogonal Procrustes Rotation) จะค้นหาไปตามเส้นตรงของระบบแกนอ้างอิง (Reference System) ระหว่างผู้สอบทั้งสองกลุ่ม

ขั้นตอนที่ 3. วิธีการที่ 1) และ 2) นำไปสู่การหมุนแกนการแปลงคะแนน โดยกำหนดตำแหน่งสเกลที่เหมาะสม (Scaling Indeterminacy) โดยยึดแนวคิดเชิงเส้นตรงของการเชื่อมโยงคะแนนตามโมเดล UIRT เช่น วิธี Mean & Sigma หรือ วิธี Stocking & Lord (1983) เป็นต้น

วิธีการของ Hirsch สามารถให้ค่าขั้นแรกโดยมีความพยายามให้สอดคล้องกับการเชื่อมโยงคะแนนตามโมเดล MIRT แต่วิธีการดังกล่าวมีหลายขั้นตอนและมีความซับซ้อนเพื่อที่จะนำไปสู่เมตริกซ์การหมุนแกนในเวคเตอร์ที่มีความแตกต่างกันระหว่างเมตริกซ์ของกลุ่มฐานและเมตริกซ์ของกลุ่มปรับเทียบคะแนน อย่างไรก็ตามในปัจจุบันนิยมนำโปรแกรม เข้ามาช่วยในกระบวนการดังกล่าวได้ง่ายและมีความสะดวกมากยิ่งขึ้น เช่น TESTFACT (Wilson, Wood, & Gibbons, 1991) และโปรแกรม NOHARM เป็นต้น โดยมีการให้ค่ามิติความสามารถของการหมุนแกนแบบตั้งฉาก ซึ่งในขั้นตอนที่ 2 (เมตริกซ์การหมุนแกนแบบตั้งฉาก (Orthogonal Procrustes Rotation) จะค้นหาไปตามเส้นตรงของระบบแกนอ้างอิง (Reference System) ระหว่างผู้สอบทั้งสองกลุ่ม) จะไม่นำมาศึกษาในโปรแกรมสำหรับผู้สอบหลายกลุ่มหรือแบบสอบหลายฉบับ ดังนั้นจึงเหลือเพียงขั้นตอนที่ 1 และ 3 ซึ่งมีวิธีการที่คล้ายคลึงกับวิธีการเชื่อมโยงคะแนนของ Li & Lissitz (2000)

2. วิธีการของ Thompson, Nering, & Davy (Thompson, Nering, & Davy's Method)

ในปี ค.ศ. 1997 Thomson et al. ได้พัฒนากระบวนการเชื่อมโยงคะแนนสำหรับแบบสอบที่มีความหลากหลายเมื่อไม่ใช้แบบสอบร่วมและกลุ่มผู้สอบที่แตกต่างกัน โดยไม่มีกลุ่มผู้สอบร่วม โดยโต้แย้งว่าสารสนเทศที่ได้จากการเชื่อมโยงคะแนนจะมีความแตกต่างกันในแบบสอบแต่ละฉบับ โดยมาจากสมมติฐานของการสุ่มผู้สอบที่มีความเท่าเทียมกันและการระบุกลุ่มเนื้อหาของข้อสอบเดียวกัน แม้ว่าวิธีดังกล่าวจะมีผลในทางปฏิบัติ เนื่องจากมีการผ่อนคลายข้อตกลงเบื้องต้นของการปรับเทียบคะแนนที่ไม่ต้องมีข้อสอบร่วมหรือผู้สอบร่วม แต่สมมติฐานดังกล่าวยังไม่เป็นที่ยอมรับ เนื่องจากสมมติฐานของความเท่าเทียมกันแบบสุ่มของกลุ่มผู้สอบค่อนข้างหายากและจำเป็นต้องอาศัยกลุ่มตัวอย่างที่มีขนาดใหญ่และเกิดจากกระบวนการสุ่ม และการระบุความคล้ายคลึงกันของกลุ่มข้อสอบค่อนข้างที่จะเป็นปัญหาและจำเป็นต้องใช้จำนวนชุดข้อสอบเป็นจำนวนมาก ซึ่ง Li & Lissitz (2000)

ให้ข้อสังเกตว่า วิธีดังกล่าวเหมาะสมในเชิงทดลองมากกว่าการนำไปใช้ปฏิบัติได้จริง ซึ่งจำเป็นที่ต้องพัฒนากระบวนการและวิธีการดังกล่าวให้ชัดเจนมากขึ้นและมีผลในทางปฏิบัติ

3. วิธีการของ Oshima, Davy, & Lee (Oshima, Davy, & Lee's Method)

Oshima, Davy, & Lee (2000) ได้พัฒนาวิธีการเชื่อมโยงคะแนนที่เรียกว่า วิธีการ "OLD" (Oshima, Davy, & Lee's Method) ซึ่งใช้ข้อสอบร่วมโดยชุดของข้อสอบร่วมจะรวมอยู่ในแบบสอบที่มีหลายฉบับเพื่อนำไปสู่สเกลร่วมกัน การแปลงคะแนนของค่าพารามิเตอร์ที่เป็นโมเดลที่ความสามารถในมิติหนึ่งสามารถชดเชยกันได้ (Compemsatory Multidimensional Model) สมการ $a_i'\theta_i + d_i$ เป็นส่วนสำคัญที่จะนำไปสู่สมการการเชื่อมโยงคะแนน ดังต่อไปนี้

$$a_i^* = (A^{-1})' a_i \quad \text{.....(2.48)}$$

$$d_i^* = d_i - a_i^* A^{-1}\beta \quad \text{.....(2.49)}$$

$$\theta_j^* = A\theta_j + \beta \quad \text{.....(2.50)}$$

โดย A (mxm) แทน เมตริกซ์การหมุนแกน (Rotation Matrix) (เมื่อ m คือจำนวนของมิติ)

β (mx1) แทน เวคเตอร์การแปลงคะแนน (Translation Vector)

* แทน ตัวบ่งชี้ค่าพารามิเตอร์การแปลงคะแนน

จากสมการจะเห็นว่าเมตริกซ์การหมุนแกน A มี 2 ฟังก์ชัน คือ a เป็นการหมุนแกนเพื่อหาทิศทางที่เหมาะสม และ b เป็นการปรับเทียบความแปรปรวนของมิติความสามารถ สำหรับเวคเตอร์การแปลงคะแนน B ใช้ในการเลื่อนจุดขึ้นจากจุดกำเนิดเพื่อเปลี่ยนตำแหน่งจุดกำเนิดใหม่ของสเกล ซึ่งความเท่าเทียมกันของการแปลงค่าพารามิเตอร์และเปลี่ยนจุดกำเนิด สามารถแสดงได้ดังสมการ

$$a_i^{*'}\theta_j^* + d_i^* = \left(a_i' A^{-1}\right)\left(A\theta_j + \beta\right) + \left(d_i - a_i' A^{-1}\beta\right) = a_i'\theta_j + d_i \quad \text{.....(2.51)}$$

จากผลที่ได้ตามสมการ ซึ่งเป็นส่วนประกอบของการแปลงคะแนน เป็นค่าทางสถิติไปยังแบบสอบอีกฉบับ ซึ่งเป็นความสัมพันธ์เริ่มแรกโดยไม่มีการเปลี่ยนแปลงความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง

นอกจากนี้ Oshima et al. ได้เปรียบเทียบวิธีการเชื่อมโยงคะแนนโดยอาศัยเกณฑ์การประเมินที่ต่างกัน ผลการวิจัยได้ข้อเสนอแนะว่าวิธีการฟังก์ชันลักษณะของแบบสอบ (Test Characteristic Function; TCF) และฟังก์ชันลักษณะของข้อสอบ (Item Characteristic Function; ICF) เป็นวิธีการที่มีความคงที่มากกว่าวิธีการอื่นที่ศึกษา (Direct Method และ Equated Function Method) โดยวิธีการทั้งสองได้รับความนิยมจัดอยู่ในประเภทของการทำให้มีค่าน้อยที่สุด ซึ่งวิธีการของ TCF เป็นวิธีการแบบพหุมิติที่แผ่ขยายมาจากวิธีการของ Stocking & Lord (1983) โดยอาศัยความแตกต่างน้อยสุดระหว่างการตอบข้อสอบของแบบสอบ 2 ฉบับ ที่เป็นข้อสอบร่วม เช่น การรวมพื้นผิวการตอบ

ข้อสอบของข้อสอบร่วม ในขณะที่วิธี ICF เป็นวิธีการที่จะทำให้ผลรวมของความแตกต่างยกกำลังสองระหว่างพื้นผิวการตอบข้อสอบมีค่าน้อยที่สุด และสุดท้ายสามารถสรุปได้ว่า วิธี TCF เป็นการประมาณค่าเมตริกซ์การหมุนแกนที่ดีที่สุดมากกว่าวิธีแยกย่อยวิธีอื่นๆ และมีความสัมพันธ์สูงกับเวคเตอร์การแปลงคะแนน

ฟังก์ชันน้อยที่สุดสำหรับวิธี TCF สามารถแสดงได้ดังสมการ

$$T(\theta) = \sum_{i=1}^{n} P_{i(\theta)} \quad \text{.....(2.52)}$$

$$\sum_{\theta} w_{\theta}\left[T_{B}(\theta) - T_{E}^{*}(\theta)\right]^{2} \quad \text{.....(2.53)}$$

โดย T_B, T^*_E แทน ค่าที่บ่งชี้คะแนนที่คาดหวังสำหรับข้อสอบร่วมของผู้สอบสำหรับแบบสอบฐานและแบบสอบปรับเทียบคะแนน

w_θ แทน ค่าแสดงน้ำหนักซึ่งอยู่ในตำแหน่งของมิติความสามารถ θ ซึ่งมีความสำคัญมากกว่าตำแหน่งอื่น

จากสมการดังกล่าวแม้ว่าน้ำหนักทั้งหมดมีค่าเท่ากับเขตพื้นที่ทั้งหมดแต่ผลการประมาณค่าที่ได้เป็นการประมาณค่าชนิดไม่ถ่วงน้ำหนัก นอกจากนี้วิธี OLD มีความเป็นหนึ่งเดียวเนื่องจากมีการประมาณค่า โดยใช้เมตริกซ์การหมุนแกนและเวคเตอร์การแปลงคะแนนที่สามารถวิเคราะห์ในคราวเดียวกันได้ แต่มีการหมุนแกนแบบตั้งฉากที่ไม่มีการกำหนดค่าที่แน่นอน หมายความว่าตำแหน่งเวคเตอร์ของข้อสอบในมิติทั้งก่อนและหลังการหมุนแกนไม่เหมือนกันนั่นเอง

4. วิธี Li & Lissitz (Li & Lissitz's Method)

ในปี ค.ศ. 2000 Li & Lissitz ได้พัฒนาวิธีการเชื่อมโยงคะแนนที่เรียกว่า "LL" (Li & Lissitz's Method) ซึ่งพัฒนากระบวนการเชื่อมโยงคะแนน 4 วิธีที่ต่างกัน ออกแบบโดยใช้ข้อสอบร่วม ผลที่ได้พบว่าวิธีการที่ดีที่สุดคือ การแปลงคะแนนรวม (Composite Transformation) ซึ่งประกอบด้วยขั้นตอนการเชื่อมโยงคะแนน 3 ส่วน คือ

1. เมตริกซ์การหมุนแกนจากวิธีการหมุนแกนแบบตั้งฉาก

2. เวคเตอร์การแปลงคะแนน ได้โดยวิธีการประมาณค่าน้อยที่สุดของความแตกต่างน้อยที่สุดระหว่างค่าพารามิเตอร์ความยากก่อนการแปลงคะแนนกับพารามิเตอร์ที่ได้จากการแปลงคะแนน

3. ค่าคงที่ของศูนย์กลางการเลื่อนแกน (Central Dilation) ได้จากวิธีของจุดที่ตัดกัน (Trace Method) ของผลรวมที่น้อยที่สุดของความคลาดเคลื่อนยกกำลังสอง

วิธีการ LL เป็นชุดของสมการเชื่อมโยงคะแนนไปยังโมเดลค่าพารามิเตอร์การแปลงคะแนนของ $a_i'\theta + d_i$ สามารถแสดงสมการองค์ประกอบได้ดังนี้

$$a_i^* = k a_i' T \quad \text{.....(2.54)}$$

$$d_i^* = d_i - a_i' T m \quad \text{.....(2.55)}$$

$$\theta_j^* = (1/k)\left(T^{-1}\theta_j + m\right) \quad \text{..... (2.56)}$$

โดย T แทน เมตริกซ์การหมุนแกนแบบตั้งฉาก

(mxm)

m แทน เวคเตอร์การแปลงคะแนนสำหรับ Location

(mx1)

K แทน ค่าคงที่ Central Dilation สำหรับการหน่วยการเปลี่ยนแปลง

จากสมการข้างต้นสามารถสรุปสมการก่อนและหลังการแปลงคะแนนได้ดังต่อไปนี้

$$a_i^{*}\theta_j^{*} + d_i^{*} = \left(ka_i^{'}T\right)(1/k)\left(T^{-1}\theta_j + m\right) + \left(d_i - a_i^{'}Tm\right) = a_i^{'}\theta_j + d \quad \text{.....(2.57)}$$

จะสังเกตได้ว่า สมการที่ 2.54-2.56 เป็นสมการเชิงคณิตศาสตร์เช่นเดียวกับ สมการ 2.48-2.50 ยกเว้น Pre-Multiplication และ Post-Multiplication ของเมตริกซ์การหมุนแกนและค่าคงที่การเลื่อนแกน

Li & Lissitz ได้เสนอวิธีการเชื่อมโยงคะแนนให้เกิดความยุติธรรมใน 3 ส่วนประกอบดังที่ได้กล่าวมาแล้ว คือ การหมุนแกน (Rotation) การแปลงคะแนน (Translation) และการกำหนดจุดศูนย์กลางการเลื่อนแกน (Central Dilation) ซึ่งเป็นส่วนประกอบที่มีความแกร่งในการให้ข้อมูลสารสนเทศเพื่อระบุขั้นตอนของค่าการแปลงคะแนนแบบพหุมิติได้ถูกต้อง

ในขณะที่วิธีการ OLD (Oshima, Davy, & Lee's Method) เป็นวิธีที่เกี่ยวข้องกับการระบุทิศทางของมิติและหน่วยการเปลี่ยนแปลง (Unit Change) ไปยังเมตริกซ์การหมุนแกนแบบไม่ใช่มุมฉากเท่านั้น ซึ่ง Li & Lissitz ได้แผ่ขยายไปอีก 2 ส่วนประกอบ คือ การหมุนแกนแบบตั้งฉากและการศึกษาค่าคงที่ศูนย์กลางการเลื่อนแกน (Central Dilation) เมื่อศัพท์คำว่า "Central" หมายถึงหน่วยการเปลี่ยนแปลง (Unit Change) โดยมีข้อตกลงเบื้องต้นว่ามีลักษณะเป็นค่าคงที่ข้ามมิติที่ศึกษา เช่น 1 สเกลลา (*k*) สามารถที่นับสำหรับหน่วยการเปลี่ยนแปลงทั้งหมด โดยระบุว่าค่าคงที่ศูนย์กลางการเลื่อนแกน (Central Dilation) ซึ่งมีลักษณะทางคณิตศาสตร์มีผลที่ตรวจสอบได้และเป็นเหตุเป็นผลที่ถูกต้องในเชิงสัมพัทธ์

กระบวนการเชื่อมโยงคะแนนที่ดีที่สุดของวิธีการ LL คือ ค่าที่น้อยที่สุดตามฟังก์ชันการหมุนแกน (T) ค่าคงที่การเลื่อนแกน (k) และเวคเตอร์การแปลงคะแนน (m) ดังสมการต่อไปนี้

$$E_1 = kA_E T - A_B \quad \text{.....(2.58)}$$

$$tr\left(E_1^{'}E_1\right) = tr\left(kA_E T - A_B\right)^{'}\left(kA_E T - A_B\right) \quad \text{.....(2.59)}$$

$$Q = \sum_{i=1}^{n}\left(d_{iB} - d_{iB}^{*}\right)^2 \quad \text{.....(2.60)}$$

โดย tr แทน Operation Matrix ของผลรวมของส่วนประกอบที่มีลักษณะเป็นเมตริกซ์แนวทแยง (Trace Matrix)

A แทน ค่าอำนาจจำแนกของข้อสอบ

D แทน ค่าความยากของข้อสอบ

B แทน แบบสอบฐาน

E แทน แบบสอบปรับเทียบคะแนน

* แทน ตัวบ่งชี้ว่ามีการแปลงคะแนนไปยังมิติของแบบสอบฐาน

5. วิธี M (Min's Method)

วิธี M เป็นวิธีการเชื่อมโยงคะแนนที่พัฒนาโดย Min (2003) ซึ่งได้ทำการพิจารณาค่าพารามิเตอร์การเลื่อนแกน (Dilation) ที่มีลักษณะเป็นสเกลลาตามวิธี "LL" (Li & Lissitz's Method) พบจุดอ่อนที่สำคัญของวิธีการดังกล่าวว่า ไม่เพียงพอสำหรับการหดหรือการขยาย (Compressing/Dilating) สเกลการเชื่อมโยงคะแนนที่มีลักษณะหลายมิติ ซึ่งค่าการเลื่อนแกนที่เป็นสเกลลา (Scalar Dilation Parameter) สามารถขยายสเกลไปแต่ละมิติให้อยู่ในลักษณะที่เป็นรูปแบบเหมือนกันโดยตรง แต่เมื่อแยกการประมาณค่าตามโมเดล MIRT ที่มีหลากหลายรูปแบบอาจส่งผลให้สเกลมีการขยายไปยังมิติที่หลากหลายในระดับต่างกัน ทำให้มีปัญหาในการประมาณค่าสำหรับการปรับเทียบแนวตั้งที่มีหลายระดับที่ต่างกัน ดังนั้น Min จึงทำแก้ไขจุดอ่อนดังกล่าว โดยมุ่งที่จะเชื่อมโยงคะแนนในลักษณะขั้นตอนทางคณิตศาสตร์ (Algorithm) ทำการแทนค่าสเกลลาการเลื่อนแกนด้วยเมตริกซ์ (Diagonal Dilation Matrix) ทำให้ผลที่ได้มีประสิทธิภาพมากขึ้น โดยวิธี M แตกต่างจากวิธี LL คือ ยอมให้มีการเปลี่ยนแปลงหน่วยความยาว (Unique Unit Change) ในแต่ละมิติมากกว่าการเปลี่ยนแปลงค่าคงที่ในมิติทั้งหมดโดยรวม อีกทั้งแตกต่างจากวิธี TCF (วิธีการของ Oshima, Davy, & Lee) และ NOP (จะกล่าวถึงในหัวข้อต่อไป) นั่นคือ วิธี M จะยึดการหมุนแกน (Rotation) และระยะของมาตรวัด อีกทั้งมีขั้นตอนที่ต่างกัน โดยใช้การหมุนแกนแบบตั้งฉาก (Orthogonal Procrustes Rotation) โดยวิธีการดังกล่าวสามารถเขียนสมการในการแปลงคะแนน ดังต่อไปนี้

$$a_i^* = a_i^* TK \quad \text{.....(2.61)}$$

$$d_i^* = d_i - a_i^* Tm \quad \text{.....(2.62)}$$

$$\theta_j^* = K^{-1}\left(T^{-1}\theta_j + m\right) \quad \text{.....(2.63)}$$

เมื่อ K_{ij} เป็น "diagonal dilation matrix"

T แทน เมตริกซ์การหมุนแกนแบบมุมฉาก (Orthogonal Rotation Matrix) สำหรับใช้ในการกำหนดทิศทาง

M แทน เวคเตอร์การแปลงคะแนนสำหรับการกำหนดตำแหน่ง (Location)

a_i แทน เวคเตอร์ของค่าพารามิเตอร์ที่มีความสัมพันธ์กับอำนาจจำแนกของข้อสอบข้อที่ i ในแต่ละมิติที่มีจำนวน m มิติ

D แทน ค่าพารามิเตอร์ที่มีลักษณะเป็นค่าสเกลลาที่มีความสัมพันธ์กับค่าความยากของข้อสอบข้อที่ i

θ_j แทน เวคเตอร์ของค่าพารามิเตอร์ที่อธิบายความสามารถของผู้สอบคนที่ j ในมิติที่ m (m-dimension space)

หมายเหตุ: * เป็นสัญลักษณ์ที่บ่งชี้ถึงค่าที่มีการแปลงคะแนนจากแบบสอบปรับเทียบคะแนนไปยังแบบสอบฐาน

จากสมการดังกล่าวจะเห็นได้ว่า K เป็น Diagonal Dilation Matrix ซึ่งสถานการณ์ที่ทำการศึกษาในครั้งนี้ ระบุให้ $K = \begin{bmatrix} k_1 & 0 \\ 0 & k_2 \end{bmatrix}$ เมื่อ k_1 ในเมตริกซ์ K แทน ส่วนประกอบของการเลื่อนแกนในมิติที่ 1 ทำนองเดียวกัน k_2 ในเมตริกซ์ K แทน ส่วนประกอบของการเลื่อนแกนในมิติที่ 2 สำหรับส่วนประกอบอื่นที่นอกเหนือแนวทแยงกำหนดให้เป็น 0 เนื่องจากมาตรวัดจะถูกนำไปแทนที่ยังแกนของมิติที่ต้องการวัดมากกว่าส่วนที่ไม่ใช่มุมฉาก ในแต่ละคอลัมน์ของ K จะประกอบด้วยจำนวนที่มีค่าไม่เท่ากับ 0 เพื่อให้สอดคล้องกับมิติที่ต้องการศึกษาในแต่ละแกน

ความเท่าเทียมกันของส่วนประกอบในการแปลงค่าพารามิเตอร์และจุดกำเนิดโดยวิธี M สามารถแสดงได้ดังสมการ

$$a_i^* \theta_j^* + d_i^* = \left(a_i^* TK\right)\left(K^{-1}\right)\left(T^{-1}\theta_j + m\right) + \left(d_i - a_i' Tm\right) = a_i' \theta_j + d_i \quad \ldots(2.64)$$

เกณฑ์น้อยที่สุด (Minimization Criteria) เพื่อใช้เป็นส่วนประกอบในการเชื่อมโยงคะแนนโดยวิธี M ใช้เกณฑ์เดียวกับวิธี LL ดังแสดง อย่างไรก็ตามส่วนประกอบของค่าพารามิเตอร์การเลื่อนแกน (Dilation; K) ซึ่งมีลักษณะเป็นเมตริกซ์มากกว่าค่าคงที่ตามวิธีของ LL เป็นจุดที่ทั้งสองวิธีมีความแตกต่างกันอย่างชัดเจน

6. การปรับเทียบคะแนนวิธี NOP (Non-Orthogonal Procrustes Method)

Reckase & Martineau (2004) ได้เสนอแนะวิธี NOP (Non-Orthogonal Procrustes Method) เป็นวิธีการที่แปลงคะแนนจากแบบสอบปรับเทียบคะแนนไปยังมิติของแบบสอบฐาน โดยต้องขึ้นอยู่กับความสอดคล้องที่เหมาะสมมากที่สุดของมิติในแบบสอบปรับเทียบคะแนนไปยังแบบสอบฐาน ข้อได้เปรียบของวิธี NOP คือ มีการตัดเมตริกซ์การเลื่อนแกน ซึ่งมีค่าเท่ากับ K ออกจากสมการ รวมถึงค่าที่เป็นสเกลล่าด้วยเช่นกัน โดยวิธี NOP มีลักษณะที่คล้ายกับวิธี TCF (Test Characteristic Function) ที่มิได้ทำการศึกษาเมตริกซ์การเลื่อนแกน (Dilation Matrix)

จากการศึกษาของ Mulaik (1972) ที่ใช้กระบวนการหมุนแกนแบบมุมแหลม (Oblique) สามารถแสดงสมการเมตริกซ์การหมุนแกน (Rotation Matrix) ดังนี้

$$T = (A'A)^{-1}A'B \quad \ldots\ldots(2.65)$$

$$a_i^* = a_i^* T \quad \ldots\ldots(2.66)$$

$$d_i^* = d_i + a_i^* Tm \quad \ldots\ldots(2.67)$$

$$\theta_j^* = \left(\tau^{-1}\theta_j - m\right) \quad \ldots\ldots(2.68)$$

เมื่อ T แทน เมตริกซ์การหมุนแกน (Rotation Matrix) สำหรับการบอกตำแหน่ง

A แทน เมตริกซ์ของค่าพารามิเตอร์อำนาจจำแนกของแบบสอบปรับเทียบคะแนน

B แทน เมตริกซ์ของค่าพารามิเตอร์อำนาจจำแนกของแบบสอบฐาน

M แทน เวคเตอร์การแปลงคะแนนสำหรับการกำหนดตำแหน่ง (Location)

a_i แทน เวคเตอร์ของค่าอำนาจจำแนกของข้อสอบข้อที่ i ในแต่ละมิติที่มีจำนวน mมิติ

D แทน ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i

θ_j แทน เวคเตอร์ของค่าพารามิเตอร์ความสามารถของผู้สอบคนที่ j ในมิติที่ m (m-dimension space)

หมายเหตุ: * เป็นค่าพารามิเตอร์ของการแปลงคะแนน (Transform Parameter) ไปยังแบบสอบฐาน

จากวิธีการ LL และ M ได้แสดงให้เห็นว่าเวคเตอร์การแปลงคะแนนโดยใช้วิธีกำลังสองน้อยที่สุด ซึ่งเป็นค่าของ Q.Q ที่เป็นผลรวมของความแตกต่างยกกำลังสองระหว่างค่าพารามิเตอร์ความยากของแบบสอบฐานและแบบสอบปรับเทียบคะแนน สามารถแสดงได้ดังสมการ

$$Q = \sum_{i=1}^{m_i} \left(d_i - d_i^*\right)^2 \quad \text{.....(2.69)}$$

เวคเตอร์การแปลงคะแนน m ตามวิธี LL จะมีความแตกต่างตามส่วนประกอบของ m ซึ่งถูกกำหนดให้เท่ากับ 0 และสามารถคำนวณได้ในคราวเดียวกัน อย่างไรก็ตามจะเหมาะสมกับการศึกษาในมิติที่มีจำนวนน้อย แต่มักจะมีความคลาดเคลื่อนสำหรับมิติที่มากขึ้น นั่นคือถ้ามีการศึกษาในมิติที่มากขึ้นจะทำให้เกิดความคลาดเคลื่อนสูงขึ้นตามไปด้วย

ในขณะที่วิธี NOP ได้แผ่ขยายแนวคิดจากสมการที่ 2.69 โดยให้ความสำคัญกับส่วนประกอบของ m ซึ่งเป็นการหาจุดกำเนิดที่เหมาะสม (Origin Indeterminacy) เพื่อแก้จุดบกพร่องดังกล่าว สามารถแสดงได้ดังสมการ

$$Q = \sum_{i=1}^{m} \left(d_i - d_i^*\right)^2 = \left(d_i - d_i^*\right)'\left(d_i - d_i^*\right) \text{.....(2.70)}$$

แต่จากสมการที่ 3.5 และ 3.6 จะได้สมการเพิ่มเติม ดังต่อไปนี้

$$d_i^* = d_i + a_i' Tm = d_i + a_i' m \rightarrow d_i^* = d_i + A_i m \text{ .....(2.71)}$$

ดังนั้น

$$Q = \left(d_i - d_i^* - A_i m\right)'\left(d_i - d_i^* - A_i m\right) \text{.....(2.72)}$$

สามารถแสดงสมการเพิ่มเติมได้ดังนี้

$$Q = \left(d_b' d_b + d_c' d_c + m' A_{\tilde{b}}' A_{\tilde{b}} m - 2d_b' d_c - 2d_b' A_{\tilde{b}} m + 2d_c' A_{\tilde{b}} m\right) \text{.....(2.73)}$$

ซึ่ง

$$d_{c-b} \equiv d_c - d_b \text{ .....(2.74)}$$

ดังนั้น

$$Q = \left(d_b' d_b + d_c' d_c + m' A_{\tilde{b}}' A_{\tilde{b}} m - 2d_b' d_c - 2d_b' A_{\tilde{b}} m + 2d_{c-b}' A_{\tilde{b}} m\right) \text{...(2.75)}$$

จะได้

$$\frac{\partial Q}{\partial m} = \frac{\partial\left(2d'_{c-b}A_{\tilde{b}}m + m'A'_{\tilde{b}}A_{\tilde{b}}m\right)}{\partial m} \quad \text{.....(2.76)}$$

จะได้

$$\frac{\partial Q}{\partial m} = 2d'_{c-b}A_{\tilde{b}} + 2m'A'_{\tilde{b}}A_{\tilde{b}} \quad \text{.....(2.77)}$$

เมื่อกำหนดให้สมการดังกล่าวมีค่า 0 จะได้ค่า m ดังต่อไปนี้

$$2d'_{c-b}A_{\tilde{b}} + 2m'A'_{\tilde{b}}A_{\tilde{b}} = 0 \quad \text{.....(2.78)}$$

$$d'_{c-b}A_{\tilde{b}} + m'A'_{\tilde{b}}A_{\tilde{b}} = 0 \quad \text{.....(2.79)}$$

$$-m'A'_{\tilde{b}}A_{\tilde{b}} = d'_{c-b}A_{\tilde{b}} \quad \text{.....(2.80)}$$

$$m' = -d'_{c-b}A_{\tilde{b}}\,(A'_{\tilde{b}}A_{\tilde{b}})^{-1} \quad \text{.....(2.81)}$$

อย่างไรก็ตามวิธี NOP ค่อนข้างมีความซับซ้อนในเชิงกระบวนการทางคณิตศาสตร์ ซึ่งขึ้นอยู่กับจำนวนของมิติที่ใช้ในการศึกษา เมื่อมีจำนวนมิติที่มากเกินไป นั่นคือมองในบริบทของโมเดลและความคลาดเคลื่อนจากการสุ่มอาจมีผลให้การแปลงคะแนนโดยวิธี NOP เกิดข้อผิดพลาดได้เช่นกัน

จากวิธีการเชื่อมโยงคะแนนในบริบทของการปรับเทียบคะแนนตั้งแต่อดีตจนถึงปัจจุบัน สามารถสรุปวิธีการและเงื่อนไขในการศึกษาการเชื่อมโยงคะแนน ได้ดังตารางที่ 2.5

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 2.5 สรุปวิธีการและเงื่อนไขที่ใช้ในการเชื่อมโยงคะแนนตามโมเดล MIRT

ผู้พัฒนา	วิธีการ	ชนิดการเชื่อมโยงคะแนน		วิธีการหมุนแกน		ลักษณะของค่า Dilation		ส่วนประกอบของการเชื่อมโยงคะแนน			จุดเด่นของวิธีการ
		Equating	Vertical Scaling	OP	NOP	Scalar	Matrix	Rotation Matrix	Translation Vector	Central Dilation	
1. Hirsch (1989)	Hirsch	√		√		√		√	√		เป็นพื้นฐานสำคัญของ MIRT Equating
2. Thompson, Nering & Davy (1997)	Thompson et al.	√	-	-	-	-		-	-	-	การออกแบบไม่ใช้แบบสอบร่วมหรือกลุ่มผู้สอบร่วม
3. Oshima, Davy & Lee (2000)	OLD	√	-	-	√	-	-	√	√	-	ประมาณค่าโดยใช้ Rotation Matrix และTransformation Vector ที่สามารถวิเคราะห์ในคราวเดียวกันได้
4. Li & Lissitz (2000)	LL	√	-	√	-	√	-	√	√	√	พัฒนาส่วนประกอบของการเชื่อมโยงคะแนนแบบพหุมิติได้อย่างมีประสิทธิภาพ
5. Min (2003)	M	√	√	√	-	-	√	√	√	√	แก้ไขจุดอ่อนของวิธี LL โดยทำการแทนค่าสเกลลาของการเลื่อนแกน ด้วยเมตริกซ์ (Diagonal Dilation Matrix) ทำให้ผลที่ได้มีประสิทธิภาพมากยิ่งขึ้น
6. Reckase & Martineau (2004)	NOP	√	√	-	√	-	-	√	√	-	มีการตัดเมตริกซ์การเลื่อนแกน (Dilation) ซึ่งมีค่าเท่ากับ K ออกจากสมการ รวมถึงค่าที่เป็นสเกลลาด้วยเช่นกัน

หมายเหตุ: OP หมายถึง การหมุนแกนแบบตั้งฉาก (Orthogonal Procrustes)

NOP หมายถึง การหมุนแกนแบบไม่ตั้งฉาก (Non-Orthogonal Procrustes)

กล่าวโดยสรุป จะเห็นว่าวิธีการเชื่อมโยงคะแนนตามโมเดล MIRT มีการพัฒนาอย่างต่อเนื่องตั้งแต่อดีตถึงปัจจุบัน ศึกษาในรูปแบบของการปรับเทียบคะแนน ซึ่งในปัจจุบันนำมาประยุกต์ใช้กับการสร้างสเกลแนวตั้งเพิ่มขึ้น เนื่องจากสอดคล้องกับสภาพจริงในสถานการณ์การสอบ อย่างไรก็ตามยังเป็นการอาศัยหลักของการปรับเทียบคะแนนอยู่เช่นเดิม

จากข้อสรุปของวิธีการเชื่อมโยงดังกล่าว พบว่ามีประเด็นสำคัญประมาณหนึ่งของการเชื่อมโยงคะแนนตามโมเดล MIRT คือ โครงสร้างมิติความสามารถที่ทำการศึกษา ดังมีรายละเอียดต่อไปนี้

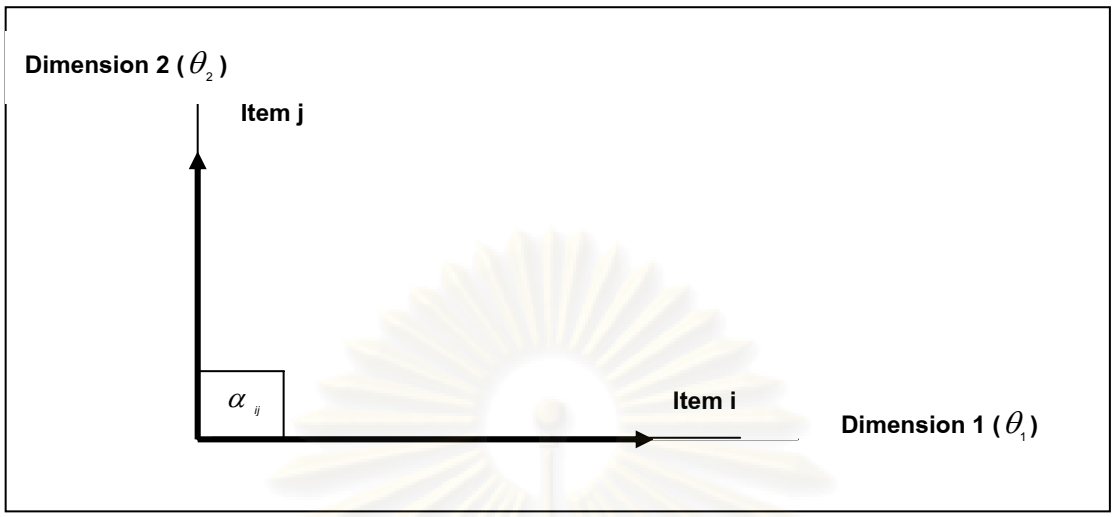
2.8 โครงสร้างมิติความสามารถ (Dimensional Structure)

แบบสอบต่าง ๆ ทั้งด้านการศึกษาและจิตวิทยาโดยทั่วไปมีลักษณะหลายมิติ นั่นคือแบบสอบมุ่งวัดโครงสร้าง (Construct) หรือคุณลักษณะแฝงมากกว่า 1 มิติ ซึ่งโครงสร้าง (Construct) หมายถึง ความเป็นตัวแทนทางทฤษฎีภายใต้ ลักษณะที่มุ่งวัด (Trait) แนวความคิด (Concept) คุณลักษณะ (Attribute) กระบวนการ (Process) และโครงสร้าง (Structure) โดยใช้แบบสอบเป็นเครื่องมือวัดคุณลักษณะแฝงดังกล่าว (Messick, 1989) ถ้าข้อสอบในชุดของแบบสอบมีลักษณะเป็นองค์ประกอบเชิงเดี่ยว (Factorial Simple) ที่วัดเพียงมิติเดียว (McDonald, 1999) เช่น แบบสอบทางคณิตศาสตร์ที่วัดโครงสร้างด้านพีชคณิต (Algebra) เพียงมิติเดียว อย่างไรก็ตามสำหรับสถานการณ์จริงของแบบสอบที่ใช้โดยส่วนใหญ่ค่อนข้างเป็นการวัดที่ซับซ้อน และมีลักษณะเป็นองค์ประกอบเชิงซ้อน (Factorial Complex) ภายใต้การวัดหลายมิติ (McDonald, 1999) เช่น แบบสอบคณิตศาสตร์ชุดหนึ่งอาจจะประกอบไปด้วยการวัดพีชคณิต (Algebra) และเรขาคณิต (Geometry) ซึ่งในกรณีนี้ชุดของแบบสอบย่อยมิติแรกกำหนดให้เป็นการวัดด้านพีชคณิต ส่วนมิติที่ 2 เป็นการวัดด้านเรขาคณิต ซึ่งเป็นการศึกษาในลักษณะ 2 มิติ โดย Zhang & Stout (1999) ได้ระบุว่า รูปแบบที่แน่นอนของการแบ่งแยกกลุ่มข้อสอบนั้นเมื่อนำแบบสอบย่อยมารวมกัน (Test Composite) ควรมีความเป็นตัวแทนกับกำหนดลักษณะของแบบสอบ (Test Specification) โดยแนะนำให้มีการ**ศึกษาโครงสร้างมิติความสามารถ** (**Dimensional Structure)** ที่มีลักษณะหลายมิติ

โดยทั่วไปโครงสร้างมิติความสามารถประกอบด้วย 3 ลักษณะ คือ โครงสร้างอย่างง่าย (Simple Structure) โครงสร้างที่ไม่ซับซ้อน (Approximate Simple Structure; APSS) และโครงสร้างที่ซับซ้อนมากขึ้น (Mixed Structure; MS) โดยมีรายละเอียดของของแต่ละลักษณะ ดังนี้

2.8.1 โครงสร้างอย่างง่าย (Simple Structure; SS)

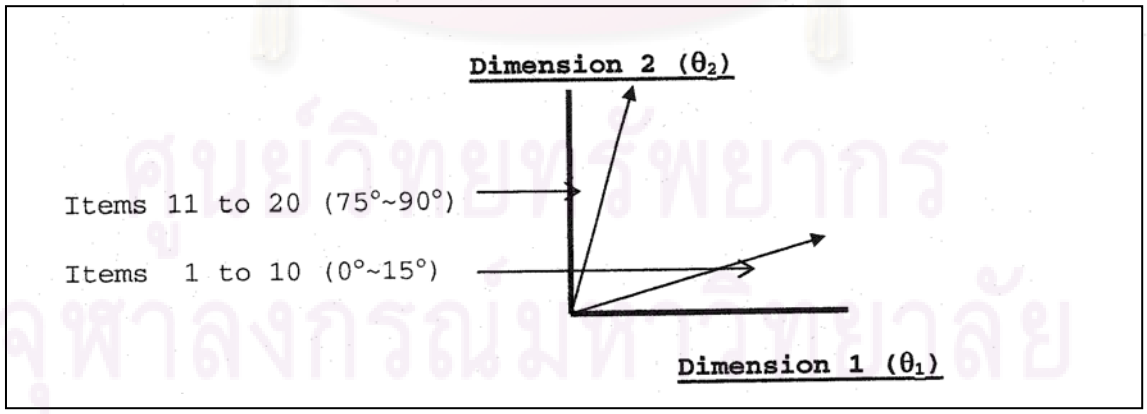
โครงสร้างอย่างง่าย (SS) เป็นโครงสร้างที่มีลักษณะของชุดข้อสอบที่สามารถแบ่งความแตกต่างของข้อสอบย่อยตามมิติใดมิติหนึ่งเพียงมิติเดียว (ซึ่งมุมระหว่างเวคเตอร์ของข้อสอบและมิติที่ศึกษาทำมุม O^0 หรือ 90^0 อย่างไรก็ตามโครงสร้างดังกล่าวมักไม่สอดคล้องกับสภาพจริง เนื่องจากมีความเป็นไปได้น้อยมากที่ชุดข้อสอบจะขึ้นอยู่กับมิติใดมิติหนึ่งเพียงมิติเดียวเท่านั้น ดังนั้นโครงสร้างดังกล่าวจึงไม่เป็นที่นิยมในการนำมาศึกษา ซึ่งสามารถแสดงได้ดังแผนภาพ



แผนภาพที่ 2.11 แบบสอบ 2 มิติที่มีลักษณะโครงสร้างมิติความสามารถอย่างง่าย (Simple Structure; SS)

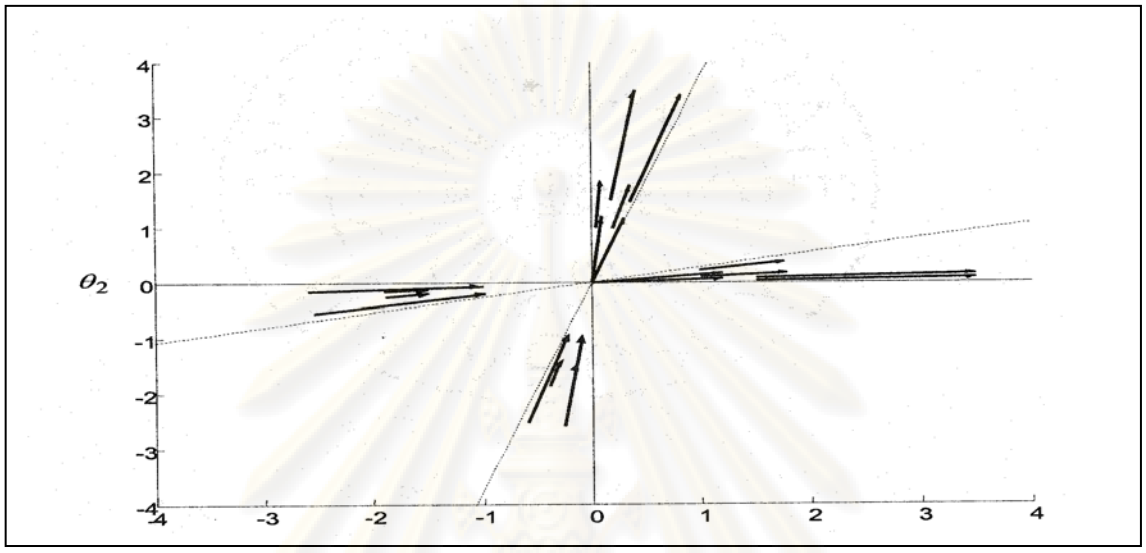
2.8.2 โครงสร้างที่ไม่ซับซ้อน (Approximate Simple Structure; APSS)

โครงสร้างที่ไม่ซับซ้อน (APSS) เป็นลักษณะโครงสร้างที่แต่ละข้อของแบบสอบมีความสัมพันธ์กันสูงในมิติที่หนึ่งมากกว่ามิติที่สอง นั่นคือ กลุ่มของข้อสอบจำนวนหนึ่งมีค่าอำนาจจำแนกที่สูงในมิติเดียวกันแต่มีค่าอำนาจจำแนกต่ำสำหรับอีกมิติหนึ่ง การศึกษาครั้งนี้ผู้วิจัยนำเสนอแบบสอบที่มีลักษณะ 2 มิติ โดยมีข้อสอบร่วมในแต่ละระดับทั้งหมด 20 ข้อ โดยในแต่ละมิติมีจำนวนข้อสอบเท่ากันมิติละ 10 ข้อ โดยลักษณะเวคเตอร์ของข้อสอบในมิติที่ 1 (Θ_1) จะทำมุมกับแกนของมิติที่ 1 ตั้งแต่ 0°–15° ($0^\circ \leq \alpha_k \leq -15^\circ$) เมื่อ α_k เป็นมุมระหว่างเวคเตอร์ในมิติที่ 1 กับแกนของมิติที่ 1 ในขณะที่ลักษณะเวคเตอร์ของข้อสอบในมิติที่ 2 (Θ_2) จะทำมุมกับแกนของมิติที่ 1 ช่วงมุมตั้งแต่ 75°–90° (90-α_k) ซึ่งโครงสร้างดังกล่าวสามารถแสดงได้ดังแผนภาพ



แผนภาพที่ 2.12 แบบสอบ 2 มิติที่มีลักษณะโครงสร้างที่ไม่ซับซ้อน (Approximate Simple Structure; APSS)

จากโครงสร้างที่ไม่ซับซ้อน แสดงให้เห็นว่า แบบสอบแต่ละชุดมีข้อสอบ 10 ข้อ แบบสอบชุดแรกมีลักษณะอยู่ในมิติที่ 1 มากกว่ามิติที่ 2 ในขณะที่ข้อสอบในชุดที่ 2 ใกล้เคียงกับมิติที่ 2 มากกว่ามิติที่ 1 โดยทิศทางของเวคเตอร์ของข้อสอบสามารถนำเสนอตัวอย่างได้ จากการศึกษาของ Min ในปี ค.ศ. 2003 ดังแผนภาพ



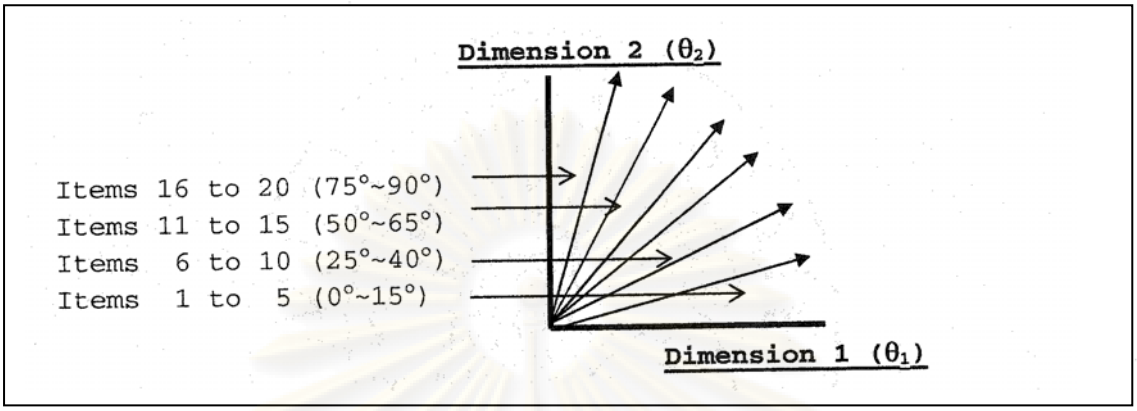
แผนภาพที่ 2.13 เวคเตอร์ข้อสอบ (Item Vectors) ที่มีลักษณะโครงสร้างที่ไม่ซับซ้อน (Approximate Simple Structure; APSS) (อ้างใน Min, 2003)

จากแผนภาพความยาวของเวคเตอร์บ่งชี้ถึงระดับของ MDISC และระยะทางระหว่างจุดกำเนิดและจุดเริ่มต้นของเวคเตอร์ (ลูกศรของเวคเตอร์ใน Quadrant ที่ 3) คือค่า MDIFF ซึ่งเวคเตอร์ทั้งหมดมุ่งสู่จุดกำเนิดซึ่งอยู่ใน Quadrant ที่ 1 และ 3 ทำให้ได้ค่าพารามิเตอร์อำนาจจำแนกมีค่าบวก (Ackerman, 1996; Reckase & McKinley, 1991 อ้างใน Min, 2003)

2.8.3 โครงสร้างที่ซับซ้อนมากขึ้น (Mixed Structure; MS)

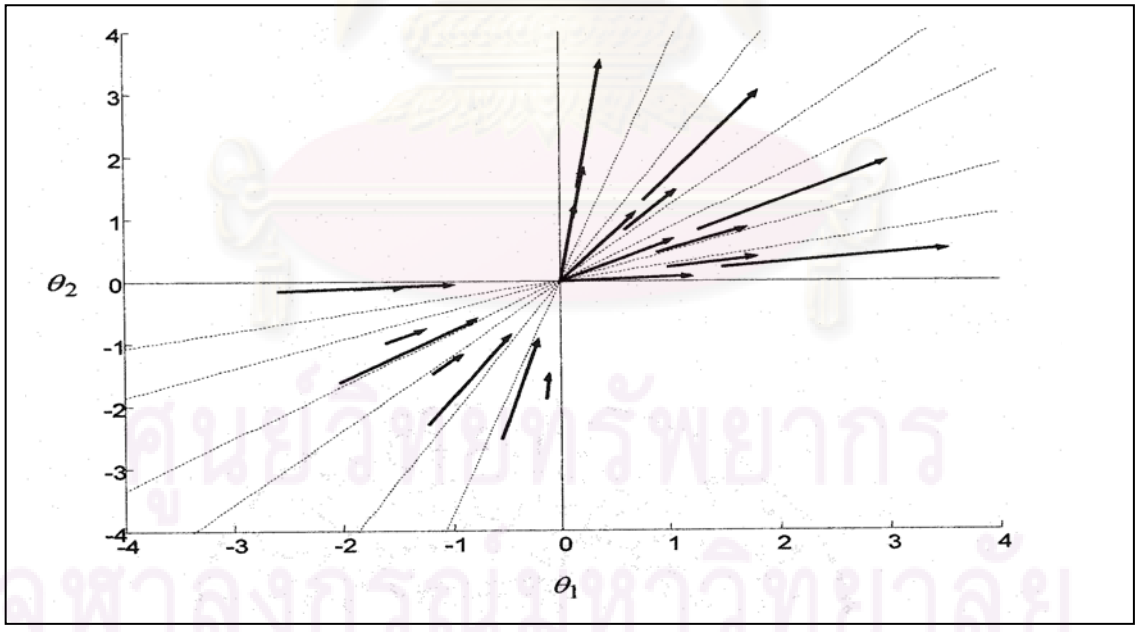
โครงสร้างที่ซับซ้อนมากขึ้น (MS) เป็นโครงสร้างที่สอดคล้องกับสถานการณ์จริงมากกว่าโครงสร้างอย่างง่าย (SS) เนื่องจากโครงสร้างโดยทั่วไปไม่ได้มีเพียงมิติใดมิติหนึ่งเท่านั้น แต่ประกอบด้วยหลายมิติ โดยลักษณะโครงสร้างจะมีการวัดในแต่ละมิติและมีการวัดมิติโดยภาพรวม (Kim, 1994) ซึ่งมีลักษณะคล้ายกรณีของโครงสร้างที่ไม่ซับซ้อน (APSS) แต่ขนาดของมุมระหว่างข้อสอบกับแกนของมิติจะต่างกันไป โดยแบบสอบย่อย 2 ชุดแรก ชุดละ 5 ข้อ จะแยกเป็นมิติที่ 1 (0°–15°) ประกอบไปข้อที่ 1-5 และ มิติที่ 2 (75°–90°) ประกอบด้วยข้อที่ 16-20 ส่วนแบบสอบย่อยอีก 2 ส่วนที่เหลือ คือ ข้อที่ 6-10 และ 11-15 จะวัดในลักษณะของมิติรวม โดยข้อที่ 6-10 มีลักษณะใกล้เคียงกับมิติที่ 1 มากกว่ามิติที่ 2 ในขณะที่ ข้อที่ 11-15 มีลักษณะใกล้เคียงกับมิติที่ 2 มากกว่ามิติที่ 1 โดยการศึกษาครั้งนี้ได้นำแนวคิดของ Wang (1986) ที่นำเสนอแนวคิดแกนโดยรวม (Reference Composite) ในลักษณะของการศึกษาแบบพหุมิติ (Multivariate Space) ค่าที่ได้สามารถสะท้อนถึงค่าอำนาจจำแนกโดยรวม (MDISCs) และค่าความแปรปรวนความแปรปรวนร่วม

ของการกระจายความสามารถที่มีลักษณะแบบหลายมิติ (Ackerman, 1992) ซึ่งโครงสร้างดังกล่าวสามารถแสดงได้ดังแผนภาพ



แผนภาพที่ 2.14 แบบสอบ 2 มิติที่มีลักษณะโครงสร้างที่ซับซ้อนมากขึ้น (Mixed Structure; MS)

จากโครงสร้างมิติความสามารถดังแผนภาพ เป็นตัวอย่างของแบบสอบย่อยที่มีทั้งหมด 4 ชุด มี 1 ชุดใกล้เคียงกับมิติที่ 1 อีก 1 ชุด ใกล้เคียงกับมิติที่ 2 ส่วนแบบสอบอีก 2 ชุดที่เหลือ จะศึกษาในลักษณะของมิติโดยรวม ซึ่งทิศทางของเวคเตอร์ของข้อสอบสามารถนำเสนอตัวอย่างได้จากการศึกษาของ Min (2003) ดังแผนภาพ



แผนภาพที่ 2.15 เวคเตอร์ข้อสอบ (Item Vectors) ที่มีลักษณะโครงที่ซับซ้อนมากขึ้น (Mixed Structure; MS) (อ้างใน Min, 2003)

จากแผนภาพมีลักษณะคล้ายโครงสร้างที่ไม่ซับซ้อน (APSS) นั่นคือ ความยาวของเวคเตอร์บ่งชี้ถึงระดับของ MDISC และระยะทางระหว่างจุดกำเนิดและจุดเริ่มต้นของเวคเตอร์

(ลูกศรของเวคเตอร์ใน Quadrant ที่ 3) คือค่า MDIFF ซึ่งเวคเตอร์ทั้งหมดมุ่งสู่จุดกำเนิดซึ่งอยู่ใน Quadrant ที่ 1 และ 3 ทำให้ได้ค่าพารามิเตอร์อำนาจจำแนกมีค่าบวก (Ackerman, 1996; Reckase & McKinley, 1991 อ้างใน Min, 2003)

กล่าวโดยสรุป โครงสร้างมิติความสามารถจะประกอบด้วย 3 ลักษณะ คือ โครงสร้างอย่างง่าย (Simple Structure; SS) โครงสร้างที่ไม่ซับซ้อน (Approximate Simple Structure; APSS) และโครงสร้างที่ซับซ้อนมากขึ้น (Mixed Structure; MS) แต่สถานการณ์โดยทั่วไปนิยมศึกษาโครงสร้างเพียง 2 ลักษณะ คือ โครงสร้างที่ไม่ซับซ้อน และโครงสร้างที่ซับซ้อนมากขึ้น เนื่องจากสอดคล้องกับสถานการณ์จริงมากกว่าโครงสร้างอย่างง่าย (Roussos, Stout & Marden, 1998; Min, 2003; Yon, 2006)

2.9 เกณฑ์การเชื่อมโยงคะแนน (Linking Criteria)

กระบวนการเชื่อมโยงคะแนน นอกเหนือจากการเลือกแบบแผนการรวบรวมข้อมูล การกำหนดนิยามและวิธีการสำหรับประมาณความสัมพันธ์ของการเชื่อมโยงคะแนนแล้ว ควรประเมินผลการเชื่อมโยงคะแนนที่ได้ ซึ่งในที่นี้จะกล่าวถึงเกณฑ์การเชื่อมโยงคะแนนในบริบทของการเชื่อมโยงคะแนน โดย American Psychological Association (Harris & Crouse, 1993 อ้างใน สุจินดา จุ้ยม่วงศรี, 2546) ได้กำหนดไว้ในมาตรฐานที่ 4.6 ว่า “ให้ระบุการตรวจสอบความเพียงพอของการเชื่อมโยงคะแนน ซึ่ง Harris & Crouse (1993) ได้รวบรวมเกณฑ์การเชื่อมโยงคะแนนที่ใช้ในงานวิจัยต่างๆ รวมทั้ง Kolen & Brennan (2004) ได้กล่าวถึงคุณสมบัติต่างๆ ของการเชื่อมโยงคะแนนที่สามารถใช้เป็นเกณฑ์การเชื่อมโยงคะแนนได้ (อ้างใน สุจินดา จุ้ยม่วงศรี, 2546) โดยมีรายละเอียดดังต่อไปนี้

2.9.1 ความเสมอภาคและเท่าเทียมกันของคะแนนที่คาดหวัง (Weak Equity or Equivalent Expected Score)

Divgi (1981) & Yen (1983) ได้เสนอกรณีพิเศษของ Lord’s Equity เรียกว่า Weak Equity จากการที่ Lord (1980) ได้กล่าวถึงคุณสมบัติ Equity ไว้ว่า “การแจกแจงคะแนนจริงแบบมีเงื่อนไขของแบบสอบแต่ละฉบับหลังจากการเชื่อมโยงคะแนนแล้วต้องเท่ากัน” ซึ่งเป็น First Order Equity ความหมายคือ ค่าคาดหวังแบบมีเงื่อนไขของคะแนนจริงที่เทียบเท่าฉบับ X มีค่าเท่ากับค่าคาดหวังแบบมีเงื่อนไขของคะแนนจริงฉบับ X ดังสมการต่อไปนี้

$$E = \left[e(Y)|\gamma(\tau)\right] = E(X|\tau) = \tau \quad \text{.....(2.82)}$$

เมื่อ τ แทน คะแนนจริงของผู้สอบจากแบสอบฉบับเก่า (X)

$(\gamma)\tau$ แทน ฟังก์ชันของ จากแบบสอบฉบับใหม่ (Y)

X แทน คะแนนฉบับเก่า

$e(Y)$ แทน คะแนนของแบบสอบฉบับใหม่ที่เทียบเท่าฉบับเก่า

ดังนั้นการเชื่อมโยงคะแนนจริงจึงมีความสัมพันธ์กับคุณสมบัติของ Equity

2.9.2 ดัชนี (Indices)

ใช้เพื่อเปรียบเทียบคะแนนแปลงที่ได้จากการเชื่อมโยงคะแนน 2 วิธี ดัชนีนี้ใช้สรุปภาพรวมของความแตกต่าง ดัชนีเดิมใช้ประเมินความคลาดเคลื่อนทางสถิติ แต่นำมาใช้กับการเชื่อมโยงคะแนนอาจจะไม่ชัดเจนนักแต่สามารถประยุกต์ใช้ดัชนีเหล่านี้โดยให้ค่าที่แท้จริง (True Value) เป็นเกณฑ์ ดัชนีกลุ่มนี้จำแนกเป็นดัชนีถ่วงน้ำหนักและดัชนีไม่ถ่วงน้ำหนัก โดยมีรายละเอียดดังนี้

(1) ดัชนีถ่วงน้ำหนักด้วยความถี่ของคะแนนดิบของแบบสอบฉบับเป้าหมาย ประกอบด้วย

(1.1) $RMSD = \left(\frac{\sum_i f_i (A_i - B_i)^2}{\sum_i f_i} \right)^{1/2}$ .....(2.83)

(1.2) $MAD = \left(\frac{\sum_i f_i |A_i - B_i|}{\sum_i f_i} \right)$ .....(2.84)

(1.3) $MAD = \left(\frac{\sum_i f_i (A_i - B_i)}{\sum_i f_i} \right)$ .....(2.85)

เมื่อ A_i แทน คะแนนสมมูล i จากแบบสอบฉบับใหม่ที่เทียบคะแนนไปยังสเกลของฉบับเก่าด้วยการเชื่อมโยงคะแนนวิธีที่หนึ่ง

B_i แทน คะแนนเกณฑ์ ซึ่งเป็นคะแนนที่เป็นจริง หรือคะแนนสมมูล I จากแบบสอบฉบับใหม่ที่ปรับเทียบคะแนนไปยังสเกลของฉบับเก่าด้วยการเชื่อมโยงคะแนนวิธีอื่น

(2) ดัชนีไม่ถ่วงน้ำหนัก เป็นการพิจารณาความแตกต่างที่เกิดขึ้นตลอดช่วงคะแนนของมาตรวัดโดยไม่กำหนดน้ำหนักในแต่ละระดับคะแนน ประกอบด้วย

(2.1) $RMSD = \left(\frac{\sum_i f_i (A_i - B_i)^2}{k} \right)^{1/2}$ .....(2.86)

(2.2) $MAD_u = \left(\frac{\sum_i f_i |A_i - B_i|}{k} \right)$ .....(2.87)

(2.3) $MSD_u = \left(\frac{\sum_i f_i (A_i - B_i)}{k} \right)$ .....(2.88)

เมื่อ K คือ จำนวนของระดับคะแนน (Score Point)

นอกจากนี้ ดัชนี RMSD ยังสามารถนำมาแยกเป็นความแปรปรวนของความแตกต่างและความลำเอียงยกกำลังสอง คือ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

$$\sum f_i (d_i)^2 / n = \sum f_i \left(d_i - \bar{d}\right)^2 / n + \bar{d}^2 \quad \text{.....(2.89)}$$

เมื่อ d_i คือ $(A_i - B_i)$

$\bar{d}$ คือ ความแตกต่างเฉลี่ย

ความลำเอียง (BIAS) ในการเชื่อมโยงคะแนน คือ ความคลาดเคลื่อนเชิงระบบ (Systematic Error)

การใช้ดัชนีเหล่านี้เป็นเกณฑ์ในการเชื่อมโยงคะแนนสามารถแปลความหมายตามขนาดดัชนี ซึ่งเป็นค่าที่เรียงลำดับได้ว่า ดัชนีที่มีค่าน้อยแสดงว่าความแตกต่างนั้นมีน้อย แต่ดัชนีขนาดเท่าใดที่ทำให้ผลการเชื่อมโยงคะแนนเป็นที่น่าพอใจยังไม่ได้มีการระบุไว้ โดย BIAS จะนำมาใช้เป็นเกณฑ์ในการพิจารณาเมื่อทราบถึงจากค่าพารามิเตอร์ที่แท้จริงของข้อสอบหรือผู้สอบ

Petersen et al. (1982) ได้เสนอดัชนีความแตกต่าง (Discrepancy Index) มีสูตรในการคำนวณ คือ

$$\text{Total error} = \frac{\sum_j f_j d_j}{n S_t^2} \quad \text{.....(2.90)}$$

เมื่อ $d_j = \left(t_j - t_j'\right)$

t_j แทน คะแนนเกณฑ์ ซึ่งเป็นคะแนนแปลงจากการเชื่อมโยงคะแนนตามทฤษฎีตอบสนองข้อสอบ

t_j' แทน คะแนนแปลงจากการเชื่อมโยงคะแนนวิธีอื่น

S_t^2 แทน ความแปรปรวนของคะแนน t_j

ค่าดัชนีที่ได้มีลักษณะเป็นค่ามาตรฐาน เพราะถ่วงน้ำหนักด้วยความแปรปรวนของคะแนนเกณฑ์ สามารถนำมาเปรียบเทียบกันได้โดยตรง ถึงแม้ในสถานการณ์ที่ได้ข้อมูลต่างกันก็ตาม นอกจากนี้ Petersen et al. ยังได้กำหนดหลักเกณฑ์การประเมินความเพียงพอในการเชื่อมโยงคะแนน ซึ่งเป็นการประเมินประสิทธิภาพของการเชื่อมโยงคะแนนตามระดับการยอมรับ 5 ระดับ โดยใช้ร้อยละของส่วนเบี่ยงเบนมาตรฐานของคะแนนเกณฑ์ยกกำลังสองเป็นการเปรียบเทียบในแต่ละระดับ

นอกจากนี้ Jaeger (1981) ยังได้เสนอดัชนีที่ใช้ตรวจสอบความเพียงพอของกระบวนการเชื่อมโยงคะแนนเชิงเส้นตรง เป็นการตรวจสอบความคล้ายคลึงกันของการแจกแจงคะแนนสะสมระหว่างคะแนนแปลงกับคะแนนของแบบสอบฉบับเป้าหมายด้วยสถิติทดสอบ Kolmogorov-Sminov Two Sample Test

2.9.3 ความคลาดเคลื่อนมาตรฐานของการเชื่อมโยงคะแนน (Standard Error of Linking)

ความคลาดเคลื่อนทั่วไปจำแนกเป็นความคลาดเคลื่อนเชิงสุ่มและความคลาดเคลื่อนเชิงระบบ ความคลาดเคลื่อนเชิงสุ่มเป็นความคลาดเคลื่อนที่เกิดจากการใช้หน่วยตัวอย่างจากประชากรมาประมาณค่าความสัมพันธ์ของการเชื่อมโยงคะแนนในประชากร สำหรับความคลาดเคลื่อนเชิงระบบได้แก่ วิธีการประมาณค่าทำให้เกิดความลำเอียงในการประมาณค่าความสัมพันธ์ของการเชื่อมโยงคะแนน เช่น เทคนิคการปรับเรียบ มีประโยชน์ช่วยลดความคลาดเคลื่อนเชิงสุ่ม แต่อาจทำให้เกิดความคลาดเคลื่อนเชิงระบบ คือ ขนาดของความคลาดเคลื่อนเชิงสุ่มลดลงเมื่อหน่วยตัวอย่างมีขนาดเพิ่มขึ้น แต่ไม่ทำให้ความคลาดเคลื่อนเชิงระบบลดลง ดังนั้นความคลาดเคลื่อนมาตรฐานของการเชื่อมโยงคะแนนคือค่าประมาณความคลาดเคลื่อนอันเนื่องมาจากการสุ่มตัวอย่างนั่นเอง

ความคลาดเคลื่อนมาตรฐานของการเชื่อมโยงคะแนนเป็นเกณฑ์ที่สำคัญในการเชื่อมโยงคะแนน แต่ในการเชื่อมโยงคะแนนบางวิธี เช่น การเชื่อมโยงคะแนนสังเกตได้ตามทฤษฎีตอบสนองข้อสอบยังไม่มีสูตรเบื้องต้นในการคำนวณ ทำให้การเชื่อมโยงคะแนนบางวิธีไม่สามารถคำนวณความคลาดเคลื่อนมาตรฐานได้ การใช้เกณฑ์ความคลาดเคลื่อนมาตรฐานยังใช้ขนาดความคลาดเคลื่อนที่เล็กกว่า แสดงถึงประสิทธิภาพที่ดีกว่า แต่ยังไม่สามารถระบุได้ว่าขนาดความคลาดเคลื่อนเท่าไหร่ที่ยอมรับได้

(1) การจำลองข้อมูล (Generated Data)

เป็นเทคนิคที่มักใช้กันบ่อยๆ ในการศึกษาการเชื่อมโยงคะแนน โดยนิยามการเชื่อมโยงคะแนนที่แท้จริง (True Equating) ด้วยโมเดลทางจิตวิทยาและสร้างข้อมูลให้เหมาะสมกับโมเดล การเชื่อมโยงคะแนนที่แท้จริงจึงใช้เป็นเกณฑ์ของการเชื่อมโยงคะแนน โมเดลเฉพาะที่สร้างขึ้นอาจมีแนวโน้มที่สนับสนุนวิธีการบางวิธี แต่แนวทางนี้จะเป็นประโยชน์ถ้าจำลองได้ใกล้เคียงกับข้อมูลที่ได้จากการทดสอบจริง

(2) การเชื่อมโยงคะแนนกลับสู่แบบสอบฉบับเดิมหรือการเชื่อมโยงคะแนนลูกโซ่ (Test Equated to Itself/ Circular Equating)

การเชื่อมโยงคะแนนกลับสู่แบบสอบฉบับเดิม ไม่ว่าจะเทียบโดยตรงหรือเทียบผ่านแบบสอบอื่น แล้วปรับสู่แบบสอบเดิมทำให้มีการเชื่อมโยงคะแนนมากกว่า 1 ครั้ง เช่นในสถานการณ์ที่มีแบบสอบ 3 ฉบับ คะแนนจากฉบับ X เทียบไปสู่สเกลฉบับ Y คะแนนจากแบบสอบฉบับ Y เทียบไปสู่สเกลฉบับ Z และคะแนนจากแบบสอบฉบับ Z เทียบกันไปสู่ฉบับ X จะเห็นว่ามีการเชื่อมโยงคะแนนกลับสู่ฉบับเดิม การเชื่อมโยงคะแนนนี้จะมีความพอเพียงก็ต่อเมื่อคะแนนดิบของแบบสอบฉบับ X มีค่าเท่ากับ 1 แปลงไปสู่คะแนน 1 หรือ 2 ไปสู่ 2 แนวคิดนี้สามารถใช้กับการเชื่อมโยงคะแนนแบบทดสอบ 3 ฉบับ คือ X Y และ Z ในแบบแผนกลุ่มสุ่ม หรือแบบแผนกลุ่มไม่เท่าเทียมกัน เมื่อใช้แบบสอบร่วมระหว่าง X กับ Y, Y กับ Z และ Z กับ X

Brennan & Kolen (2004) ได้ชี้ให้เห็นประเด็นที่น่าสนใจว่า 1) การเชื่อมโยงคะแนน Identity จะให้ผลดีเมื่อเทียบคะแนนแบบลูกโซ่ และวิธีการเชื่อมโยงคะแนนที่ประมาณค่าพารามิเตอร์น้อยกว่า เช่น

การเชื่อมโยงคะแนนเชิงเส้นตรง มีความโน้มเอียงว่าจะเป็นวิธีที่ดีกว่าวิธีที่มีการประมาณค่าพารามิเตอร์มากกว่า เช่น การเชื่อมโยงคะแนนอีควิเปอร์เซนไทล์ ภายใต้แบบแผนกลุ่มไม่ทัดเทียมกันที่ใช้ข้อสอบร่วม 2) ผลของการเปรียบเทียบขึ้นอยู่กับแบบสอบที่ใช้เริ่มต้น คือ เมื่อแบบทดสอบฉบับ X เทียบกลับสู่ฉบับเดิม โดยผ่าน Y และ Z อาจแตกต่างจากผลที่ได้เมื่อ Z เทียบไปสู่ฉบับเดิม โดยผ่าน X และ Z

(3) หน่วยตัวอย่างขนาดใหญ่ (Large Sample Criterion)

ข้อมูลจากตัวอย่างผู้สอบขนาดใหญ่ใช้เป็นตัวแทนของประชากร ตัวอย่างขนาดเล็กจะถูกสุ่มออกมา และผลที่ได้จากหน่วยตัวอย่างขนาดเล็ก จะนำมาเปรียบเทียบผลการเชื่อมโยงคะแนนจากหน่วยตัวอย่างขนาดใหญ่ ซึ่งถือเป็นความสัมพันธ์ของการเชื่อมโยงคะแนนที่แท้จริงเกณฑ์นี้มีข้อจำกัดเพราะว่ากลุ่มผู้สอบขนาดใหญ่ไม่สามารถหาได้ ผลที่ได้มีความหมายเฉพาะกลุ่มผู้สอบที่มีลักษณะเดียวกัน

(4) การเปรียบเทียบความคงเส้นคงวาของผลการเชื่อมโยงคะแนน (Comparisons of the Consistency of Equating Results)

ความคงเส้นคงวาของการเชื่อมโยงคะแนน ได้มาจากการนำวิธีการเชื่อมโยงคะแนนหลายวิธีไปใช้ในการปรับเทียบในสถานการณ์อื่นๆ แล้วพิจารณาผลที่ได้ว่าแตกต่างหรือเหมือนกันกับผลที่ได้จากสถานการณ์เดิม ถ้าให้ผลเหมือนกันแสดงว่ามีความแม่นยำในการเชื่อมโยงคะแนน การใช้ความคงเส้นคงวาเป็นเกณฑ์ในการศึกษาการเชื่อมโยงคะแนนต้องระมัดระวังว่าอาจเกิดความไม่คงเส้นคงวาในสถานการณ์ที่ไม่เหมาะสมหรือไม่เพียงพอกับวิธีการดังกล่าว

(5) การสอบทานผล (Replication Samples/ Cross-Validation Studies/ Stability)

เป็นการนำผลที่ได้รับในกลุ่มตัวอย่างหนึ่งไปใช้กับกลุ่มตัวอย่างอื่น เป็นการตรวจสอบความคงที่ของการเชื่อมโยงคะแนน ซึ่งเป็นเกณฑ์ภายนอกที่ใช้เพื่อเปรียบเทียบความแตกต่างของการเชื่อมโยงคะแนนด้วยวิธีที่แตกต่างกัน คุณภาพของการวัดด้วยแบบสอบทานผลแสดงความคงที่ของการเชื่อมโยงคะแนน กลุ่มตัวอย่างที่ใช้ตรวจสอบผล เรียกว่า กลุ่มสอบทานผลมีลักษณะเช่นเดียวกับกลุ่มปรับเทียบคะแนน กลุ่มสอบทานผลจะทำแบบสอบทั้งสองฉบับ แล้วนำคะแนนจากแบบสอบฉบับหนึ่งไปแปลงจากตารางที่สร้างไว้ แล้วพิจารณาผลต่างระหว่างคะแนนที่ทำได้จริงกับคะแนนแปลงเป็นการตรวจสอบความคลาดเคลื่อนจากการเชื่อมโยงคะแนน

Kolen & Whitney (1982) ได้ใช้เกณฑ์สอบทานผลในการเปรียบเทียบความแตกต่างระหว่างวิธีการเชื่อมโยงคะแนน 4 วิธี โดยนำคะแนนของกลุ่มสอบทานผล ที่ทำแบบสอบทั้งสองฉบับ ไปตรวจสอบด้วยดัชนีความแตกต่างระหว่างการแจกแจงคะแนนแปลงกับคะแนนที่ทำได้จริงเรียกว่าดัชนีเปรียบเทียบเปอร์เซ็นไทล์ เป็นดัชนีความแตกต่างกำลังสองเฉลี่ยระหว่างคะแนนที่ได้กับคะแนนแปลง มีสูตรในการคำนวณดังนี้

$$C = \sum \frac{(X_i - X^*)}{nk} \qquad \text{.....(2.91)}$$

เมื่อ n แทน จำนวนของคะแนนดิบของกลุ่มสอบทานผล

k แทน จำนวนข้อสอบในแบบสอบร่วมที่ใช้

ค่า C ที่ได้ถ้ามีค่าน้อยหมายความว่า รูปแบบการเชื่อมโยงคะแนนที่นำมาสร้างตารางปรับเทียบคะแนนนั้นมีความเหมาะสมและเพียงพอที่จะให้ผลการแปลงคะแนนอย่างคงเส้นคงวา ดัชนีนี้เหมาะสมที่จะนำไปใช้กับกลุ่มสมมูลแบบสุ่ม 2 กลุ่ม ที่ทำแบบสอบคนละฉบับ หรือกลุ่มหนึ่งกลุ่มที่ทำแบบสอบทั้งสองฉบับ ซึ่ง Kolen & Whitney ได้กล่าวถึงเกณฑ์สอบทานผลว่าเป็นเกณฑ์ภายนอกที่เป็นองค์ประกอบสำคัญของการเปรียบเทียบความแตกต่างระหว่างการเชื่อมโยงคะแนนด้วยวิธีการที่แตกต่างกัน

(6) เกณฑ์อื่น ๆ (Other Criteria)

เกณฑ์อื่นๆ ที่เคยใช้ประเมินการเชื่อมโยงคะแนน เช่น Budescu (1985 อ้างใน สุจินดา จุ้ยม่วงศรี, 2546) ใช้ฟังก์ชันโมโนโทนิคของความสัมพันธ์ระหว่างข้อสอบร่วมกับไม่ได้ร่วมเป็นตัววัดประสิทธิภาพของการเชื่อมโยงคะแนน Patience (1981) เคยใช้ความสัมพันธ์ของคะแนนผู้สอบจากสเกลของคะแนนแปลงกับเกณฑ์ซึ่งเป็นสเกลอ้างอิง เพื่อศึกษาประสิทธิภาพสัมพัทธ์ของการปรับเทียบแนวตั้ง Cope (1992) ได้แนะนำให้ใช้เกณฑ์ภายนอกเพื่อประเมินการเชื่อมโยงคะแนน เช่น การเพิ่มความตรงเชิงพยากรณ์หรือการเปรียบเทียบอัตราสอบผ่านในการเชื่อมโยงคะแนนวิธีต่างๆ

นอกจากนี้ Kolen & Brennon (2004) ได้แนะนำให้ใช้คุณสมบัติของการเชื่อมโยงคะแนนเป็นเกณฑ์ในการประเมินผลการเชื่อมโยงคะแนน คุณสมบัติเหล่านั้น คือ

คุณสมบัติความสมมาตร (Symmetry Property) คือ ผลของการเชื่อมโยงคะแนนจากแบบสอบฉบับ X ไปยังฉบับ Y หรือการเชื่อมโยงคะแนนจากแบบสอบฉบับ Y ไปยังฉบับ X จะต้องให้ผลเหมือนกัน

คุณสมบัติโครงสร้างเนื้อหาที่คล้ายคลึงกัน (Same Specifications Property) คือ แบบสอบสองฉบับที่นำมาปรับเทียบคะแนนต้องมีความคล้ายคลึงกันของเนื้อหา ควรสร้างจากตารางวิเคราะห์หลักสูตรเดียวกัน เพื่อให้คะแนนของแบบสอบสามารถแทนที่กันได้

คุณสมบัติการเชื่อมโยงคะแนนสังเกตได้ (Observed Score Equating Properties) คือ ในการเชื่อมโยงคะแนนสังเกตได้ ลักษณะการแจกแจงคะแนนถูกกำหนดให้เท่ากันในประชากรของผู้สอบ (Angoff, 1984 อ้างใน Kolen & Brennon, 2004) เช่นคุณสมบัติของการเชื่อมโยงคะแนนอิควิเปอร์เซ็นไทล์ คะแนนแปลงของแบบสอบฉบับ X มีการแจกแจงคะแนนเหมือนกับแบบสอบฉบับ Y มีฟังก์ชันการเชื่อมโยงคะแนน คือ

$$G^{*}\left[e_{y}(X)\right]=G(y) \quad \text{.....(2.92)}$$

ภายใต้คุณสมบัติของการเชื่อมโยงคะแนนเชิงเส้นตรง ซึ่งคะแนนแปลงของแบบสอบทั้งสองฉบับมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากัน เมื่อการเชื่อมโยงคะแนนอีควิเปอร์เซ็นไทล์มีคุณสมบัติดังกล่าว การเชื่อมโยงคะแนนเชิงเส้นตรงก็มีคุณสมบัติเหมือนคะแนนสังเกตได้ด้วย

คุณสมบัติการไม่แปรเปลี่ยนกลุ่ม (Group Invariance Property) คือ คะแนนที่ปรับเทียบแล้วจะไม่แปรผัน ไม่ว่าจะปรับเทียบคะแนนจากกลุ่มใด สามารถนำผลไปใช้กับกลุ่มอื่นได้

สำหรับเกณฑ์ที่ใช้ประเมินผลการเชื่อมโยงคะแนนที่ได้จากการประเมินการปฏิบัติ การประเมินตามสภาพจริงที่ข้อสอบตรวจให้คะแนนหลายค่า ส่วนใหญ่ไม่ได้พัฒนาเกณฑ์ขึ้นมาใหม่ยังคงใช้เกณฑ์ที่ใช้กับการเชื่อมโยงคะแนนที่ข้อสอบตรวจให้คะแนนสองค่า ซึ่ง Harris & Crouse (1993) ได้กล่าวว่าอาจเป็นเพราะวิทยาการของการเชื่อมโยงคะแนนในแบบสอบประเภทนี้ยังได้รับการพัฒนาไม่มากนัก แต่การทดสอบด้วยเครื่องมือที่ใช้คำถามปลายเปิด และการประเมินการปฏิบัติได้มีการใช้สูงมากขึ้นเรื่อยๆ ในการวัดและประเมินผลในปัจจุบัน การประเมินผลการเชื่อมโยงคะแนนประเภทนี้ควรต้องมีการพัฒนาเกณฑ์ตามไปด้วย

กล่าวโดยสรุป การนำเกณฑ์ดังกล่าวไปใช้ในการประเมินผลการเชื่อมโยงคะแนน มิได้นำเกณฑ์เดียวไปใช้ได้เลย เช่น การใช้เกณฑ์กลุ่มสอบทานผลควบคู่กับดัชนีที่บ่งบอกความแตกต่างเป็นเกณฑ์การเทียบคะแนน และยังสร้างเกณฑ์ประเมินค่าดัชนีความแตกต่างด้วย ดังนั้น ไม่ควรมีการโต้เถียงว่าเกณฑ์ใดดีที่สุด แต่ควรมีการพิจารณาว่าเกณฑ์นั้นมีประโยชน์หรือจำเป็นในสถานการณ์นั้นๆ เกณฑ์การเชื่อมโยงคะแนนยังจำเป็นต่อการประเมินผล (Harris & Crouse, 1993) และการเปลี่ยนแปลงเกณฑ์จะทำให้มีการเปลี่ยนแปลงผลสรุปที่ได้ (Skaggs, 1990 อ้างใน Harris & Crouse, 1993) แต่โดยทั่วไปเมื่อศึกษาในบริบทของการเชื่อมโยงคะแนนแบบพหุมิติตามโมเดล MIRT โดยทั่วไปนิยมใช้ BIAS และ RMSE เป็นเกณฑ์ที่สำคัญเพื่อพิจารณาถึงคุณภาพของการปรับเทียบคะแนน

2.10 งานวิจัยที่เกี่ยวข้องกับการเชื่อมโยงคะแนนตามโมเดล MIRT

การเชื่อมโยงคะแนนตามโมเดล MIRT ได้มีการศึกษาเฉพาะในต่างประเทศ ส่วนในประเทศไทยยังไม่มีผู้ใดศึกษาโดยตรง ดังนั้นมีความจำเป็นที่ต้องศึกษางานวิจัยที่เกี่ยวข้องกับการนำโมเดล MIRT มาประยุกต์ใช้ในการเชื่อมโยงคะแนนที่เป็นบริบทของการปรับเทียบคะแนน เพื่อเป็นสารสนเทศสำคัญในการพัฒนาองค์ความรู้ดังกล่าวในประเทศ โดยมีรายละเอียดดังนี้

Hirsch (1989) ได้ศึกษาวิธีการเชื่อมโยงคะแนนแบบพหุมิติตามโมเดล MIRT แบบโลจิสติก 2 พารามิเตอร์ (M2PLM) ศึกษาทั้งจากการจำลองข้อมูลและการใช้ข้อมูลจริง ออกแบบการเก็บรวบรวมโดยใช้ผู้สอบร่วม สำหรับการวัดคุณลักษณะภายใน 2 มิติ ประมาณค่าพารามิเตอร์ของข้อสอบและพารามิเตอร์ความสามารถ โดยใช้โปรแกรม Multidimensional Item Response Theory Estimation (MIRTE) มีการประเมินความคลาดเคลื่อนของการปรับเทียบคะแนนโดยการเปรียบเทียบความแตกต่างของค่าเฉลี่ยและการเปรียบเทียบค่าสัมบูรณ์ความแตกต่างของค่าเฉลี่ย ระหว่างความสามารถที่แท้จริงกับความสามารถที่สังเกตได้

ผลการวิจัยพบว่าการปรับเทียบมีประสิทธิภาพ เมื่อเปรียบเทียบกับความสามารถที่แท้จริง ซึ่งเป็นเทคนิคที่ใช้ในการศึกษาครั้งนี้ อย่างไรก็ตามผลการประเมินความคงที่ (Stability) ของการประมาณค่าพารามิเตอร์ความสามารถ พบว่า ยังไม่เป็นที่น่าพอใจ

Oshima & Davey (1994) ได้ประเมินประสิทธิภาพการเชื่อมโยงคะแนนตามโมเดลที่การประมาณค่าความสามารถในมิติต่างกันสามารถชดเชยกันได้ (Compensatory Model) โดยทำการจำลองข้อมูล

ความสามารถต่างกัน 2 ชุด เพื่อศึกษาใน 2 มิติ ออกแบบวิธีการเก็บรวบรวมข้อมูลโดยใช้ชุดของข้อสอบร่วมสำหรับผู้สอบ 2 กลุ่ม คือ กลุ่มปรับเทียบคะแนนและกลุ่มฐาน ทำการเชื่อมโยงคะแนนให้อยู่ในสเกลเดียวกันโดยศึกษาจากค่าความแตกต่างของเวคเตอร์ค่าเฉลี่ยและความแตกต่างของเมตริกซ์ความแปรปรวนความแปรปรวนร่วม หลังจากมีการประมาณค่าพารามิเตอร์ของข้อสอบแล้วจะมีการกำหนดค่าเฉลี่ยของค่าพารามิเตอร์ข้อสอบเพื่อใช้ปรับเทียบคะแนนทั้งสองกลุ่มโดยใช้โปรแกรม NOHARM ประมาณค่าพารามิเตอร์ ส่วนเกณฑ์ที่ใช้ประเมินประสิทธิภาพจะศึกษาจากความคลาดเคลื่อน RMSE (Root Mean Square Error)

ผลการวิจัยพบว่า การใช้วิธีความแตกต่างของเมตริกซ์ความแปรปรวนความแปรปรวนร่วมมีประสิทธิภาพน้อยกว่าการศึกษาจากค่าความแตกต่างของเวคเตอร์ค่าเฉลี่ย และผลการวิจัยยังพบว่าการแปลงคะแนนแบบมุมแหลม จะสามารถประมาณค่าพารามิเตอร์ได้ใกล้เคียงกับค่าพารามิเตอร์ที่แท้จริงมากกว่าการแปลงคะแนนแบบมุมฉาก อย่างไรก็ตามมีข้อจำกัดในการใช้โปรแกรม NOHARM สำหรับความสัมพันธ์ θ_s ที่เป็นผลเนื่องมาจากการประมาณค่าพารามิเตอร์ข้อสอบที่มีค่าสุดโต่ง ซึ่งส่งผลกระทบต่อการเชื่อมโยงคะแนน นอกจากนี้ยังพบว่าการวิจัยนี้มีข้อจำกัดคือ มีการศึกษาสมการการสร้างสเกล (Scaling Equation) เพียง 1 สมการเท่านั้น ดังนั้นการศึกษาครั้งต่อไปควรที่จะมุ่งไปที่การสร้างสมการการสร้างสเกลหลายสมการที่ต่างกัน และจำเป็นต้องมีการตรวจสอบผลที่ได้จากการใช้โปรแกรม NOHARM ให้เหมาะสมและสอดคล้องกับโมเดลมากขึ้น

Thompson et al. (1997) ได้พยายามที่จะพัฒนาวิธีการเชื่อมโยงคะแนนตามโมเดล MIRT โดยไม่ต้องใช้ผู้สอบร่วมหรือข้อสอบร่วมภายใต้สถานการณ์การสอบที่มีแบบสอบแตกต่างกัน ซึ่งใช้การสุ่มกลุ่มผู้สอบที่มีขนาดใหญ่ อยู่ภายใต้ข้อตกลงเบื้องต้นว่า แกนของจุดกำเนิดและความสัมพันธ์ระหว่างแกน เป็นจุดร่วมกันของทั้งสองกลุ่ม โดยกลุ่มที่ได้ต้องเป็นกลุ่มสุ่มที่เท่าเทียมกัน อย่างไรก็ตามการกำหนดการหมุนแกนของแต่ละกลุ่มจะเกิดขึ้นพร้อมกับกระบวนการในการประมาณค่าพารามิเตอร์ซึ่งคล้ายคลึงกับการวิเคราะห์องค์ประกอบที่นำไปสู่การหมุนแกนในตำแหน่งที่เหมาะสม (Rotation Indeterminacy) ซึ่งเป็นวิธีที่สามารถระบุความคล้ายคลึงกันของเนื้อหาเพื่อใช้ในการจัดกลุ่ม (Cluster) ของแบบสอบต่างชุดกัน จากนั้นจะทำการหมุนแกนจนอยู่ในสเกลเดียวกัน ซึ่งวิธีในการระบุกลุ่มของมวลเนื้อหาจากแบบสอบที่แตกต่างกันจากการศึกษาของ Reckase, Thompson, & Nering (1997)

ผลการวิจัยพบว่าความสำเร็จในการปรับเทียบคะแนนขึ้นอยู่กับวิธีการที่ใช้ระบุการจัดกลุ่มมวลเนื้อหาที่คล้ายคลึงกันจากแบบสอบที่ต่างกันหลายฉบับ อย่างไรก็ตามการศึกษาครั้งนี้ยังเป็นลักษณะในเชิงทดลอง ซึ่งจำเป็นที่จะต้องศึกษาให้ประเด็นดังกล่าวให้มีความชัดเจนยิ่งขึ้น

Oshima, Davey, & Lee (1997 อ้างใน Oshima, Davey, & Lee, 2000) ได้ศึกษาวิธีการปรับเทียบคะแนนแบบพหุมิติ ใน 2 มิติ การประมาณค่าความสามารถจะอาศัยเมตริกซ์การหมุนแกนและเวคเตอร์การแปลงคะแนน โดยใช้ฟังก์ชันทางคณิตศาสตร์ให้น้อยที่สุด เช่น การระบุโดยความแตกต่างยกกำลังสองของพื้นผิวคุณลักษณะของแบบสอบทั้งสองฉบับ อย่างไรก็ตาม จากการศึกษาของ Oshima, Davey, & Lee (1997) ไม่มีการประมาณค่าพารามิเตอร์การเลื่อนแกน (Dialation) โดย

เมตริกซ์การหมุนแกนที่ศึกษาเป็นรูปแบบของความหลากหลายของการแปลงคะแนนรวม เพื่อนำมาใช้ในสถานการณ์เมตริกซ์การแปลงคะแนนแบบไม่ตั้งฉาก

Oshima, Davey, & Lee (2000) ได้ทำการศึกษาวิธีการเชื่อมโยงคะแนนในบริบทของการปรับเทียบคะแนนตามโมเดล MIRT 4 วิธี คือ 1) Direct Method 2) Equated Function Method 3) Test Characteristic Function และ 4) Item Characteristic Function ทำการประเมินประสิทธิภาพการปรับเทียบคะแนนโดยอาศัยข้อมูลจากการจำลองข้อมูล ค่าพารามิเตอร์ของข้อสอบได้มาจากการใช้ข้อมูลจริงจากแบบสอบ ACT ที่วัดความสามารถทางคณิตศาสตร์ ในปี 1992 ซึ่งเป็นแบบสอบมุ่งวัด 2 มิติ จำนวนข้อสอบ 40 ข้อ ส่วนค่าพารามิเตอร์ความสามารถของผู้สอบได้มาจากการจำลองข้อมูล โดยค่าพารามิเตอร์ของข้อสอบจะถูกประมาณค่าให้เป็นมาตรฐาน จากทั้งสองกลุ่มที่เป็นอิสระจากกัน โดยใช้โปรแกรม NOHRAM ขั้นตอนในการเชื่อมโยงคะแนนจากแบบสอบหนึ่งไปยังแบบสอบหนึ่งโดยใช้โปรแกรม Windows-based ที่เรียกว่า IPLINK เกณฑ์ที่ใช้ในการประเมินประสิทธิภาพการปรับเทียบคะแนน คือ BIAS และ RMSE

ผลการศึกษาพบว่าวิธี Test Characteristic Function (TCF) และ Item Characteristic Function (ICF) มีความคงที่มากกว่าวิธี Direct และ Equated Function นอกจากนี้พบว่าทั้งสองวิธีดังกล่าวเชื่อมโยงได้ใกล้เคียงกับค่าพารามิเตอร์มากกว่าวิธี Direct และ Equated Function อย่างไรก็ตามยังไม่สามารถตอบได้ว่าวิธีการใดดีที่สุด ต้องขึ้นอยู่กับสถานการณ์ที่เลือกใช้เป็นสำคัญ เช่น ถ้าคะแนนความสามารถที่แท้จริงของทั้งสองมีความเท่าเทียมกันวิธีที่ควรเลือกใช้ คือ TCF เนื่องจากพิจารณาความแตกต่างของคะแนนจริงให้น้อยที่สุด หรืออีกทางหนึ่งถ้าต้องการตรวจสอบการทำหน้าที่แตกต่างกันของข้อสอบวิธีที่เหมาะสมน่าจะเป็น ICF เป็นต้น เนื่องจากการศึกษาครั้งนี้มีข้อจำกัดของโปรแกรมที่ศึกษาใน 2 มิติ ในการวิจัยครั้งต่อไปควรศึกษาเงื่อนไขอื่นที่เกี่ยวข้องกับการปรับเทียบคะแนน เช่น การศึกษาจำนวนมิติให้มากขึ้น ประกอบกับพัฒนาเทคนิควิธีในการเชื่อมโยงคะแนนแบบพหุมิติและการตรวจสอบประสิทธิภาพจากเงื่อนไขดังกล่าว เป็นต้น

Li & Lissitz (2000) ได้ประเมินประสิทธิภาพของเทคนิควิธีที่ใช้เชื่อมโยงคะแนนในบริบทของการปรับเทียบคะแนนตามโมเดล MIRT แบบโลจิสติก 2 พารามิเตอร์ โดยการจำลองข้อมูลให้เหมาะกับข้อมูลที่มีลักษณะแบบพหุมิติ ซึ่งมีการแจกแจงแบบปกติพหุ (MVN(0,1)) การจำลองข้อมูลในครั้งนี้ค่าพารามิเตอร์ของข้อสอบได้มาจากแบบสอบ ACT Form 24B จากการศึกษาของ Reckase (1985) เพื่อตรวจสอบความคลาดเคลื่อนของการประมาณค่าพารามิเตอร์ โดยมีส่วนประกอบของการเชื่อมโยงคะแนน 3 ส่วน คือ 1) Orthogonal Procrustes Rotation 2) Translation Tranformation และ 3) Single Dilation เพื่อนำไปสู่การประมาณค่า 3 ค่า คือ 1) Othogonal Procruster Rotation Matrix (T) 2) Diagonal Dilation Matrix (K) และ 3) Translation Vector (m) ซึ่งเป็นส่วนประกอบสำคัญในการแปลงคะแนนองค์ประกอบเพื่อเชื่อมโยงคะแนนจากแบบสอบปรับเทียบคะแนนไปยังแบบสอบฐาน โดยมีวิธีที่ใช้ในการประมาณค่า (Estimators) ดังนี้ 1) ประมาณค่า k โดยวิธี Ratio of Eigenvalues, Trace, Least Square และ MTRs 2) ประมาณค่าพารามิเตอร์ในมิติของ m_1 และ m_2 โดยวิธี Least Square และ MTRs จำลองข้อมูลเพื่อประเมินประสิทธิภาพของการเชื่อมโยงคะแนน ซึ่งออกแบบวิธีการเก็บรวบรวม

ข้อมูลสำหรับผู้สอบต่างกลุ่มโดยใช้แบบสอบร่วม ภายใต้เงื่อนไขดังนี้ 1) เทคนิควิธีที่ใช้ในการเชื่อมโยงคะแนน (Orthogonal Procrustes Rotation, Translation Tranformation และ Single Dilation) 2) ขนาดกลุ่มตัวอย่างที่แตกต่างกัน (จำนวน 1,000, 2,000 และ 4,000 คน) 3) กระบวนการในการปรับเทียบคะแนน 4) จำนวนของข้อสอบร่วม (15 และ 25 ข้อ) ซึ่งเลือกมาจากแบบสอบ ACT Form 24B 5) สถานการณ์ในการปรับเทียบคะแนน (Horizontal Equating และ Vertical Equating) รวมเงื่อนไขในการศึกษาทั้งหมด 72 เงื่อนไข ทำการประมาณค่าพารามิเตอร์ตามโมเดล MIRT โดยใช้โปรแกรม TESTFACT (Wilson et al., 1991) และประมาณค่าพารามิเตอร์การแปลงคะแนน จากแบบสอบฉบับหนึ่งไปยังแบบสอบฉบับหนึ่ง โดยใช้โปรแกรม MDEQUATE (Li, 1996) โดยเขียนโปรแกรมใน MATLAB (Version 5.1) ซึ่งการประเมินประสิทธิภาพการปรับเทียบคะแนนตามโมเดล MIRT พิจารณาจากค่า BIAS และ RMSE

ผลการวิจัยพบว่า วิธีการที่เหมาะสมที่สุดที่จะทำให้การแปลงคะแนนค่าพารามิเตอร์มีความถูกต้องมากที่สุด คือ การรวมกันระหว่าง Procrustes Rotation, Ratio of Trace และการประมาณค่าแบบ Least Square เพื่อใช้ประมาณค่าเมตริกซ์การแปลงคะแนนแบบมุมฉาก (Rotational Transformation Matrix) พารามิเตอร์การเลื่อนแกน (Dilation Parameter) และชุดของพารามิเตอร์การเปลี่ยนผ่าน (Set of Translation Parameter) ซึ่งการประมาณค่าแบบ Procrustes Rotation เพื่อประมาณค่า Rotation Matrix พิจารณาจาก BIAS และ RMSE พบว่าอยู่ในระดับที่น่าพอใจ ส่วนการประมาณค่าแบบ Ratio of Trace เป็นวิธีการที่ดีที่สุดที่ใช้ในการประมาณค่าการเลื่อนแกน (Dilation Parameter) รองลงมาคือ Eigenvalues และ MTRs ตามลำดับ ส่วนวิธีการประมาณค่าแบบ Least Square มีความคลาดเคลื่อนน้อยกว่า MTRs สำหรับการประมาณค่า พารามิเตอร์การแปลงคะแนน โดยเฉพาะจากการศึกษาในมิติที่ 2 นอกจากนี้ผลการวิจัยพบว่าสามารถนำเทคนิคการเชื่อมโยงคะแนนตามโมเดล MIRT ไปใช้กับข้อมูลที่มีลักษณะแบบเอกมิติได้อย่างมีประสิทธิภาพ

Briggs (2003) ได้ศึกษาวิธีการวัดแบบพหุมิติ โดยใช้ Multidimensional Random Coefficient Multinomial Logit Model (MRCM) ที่แผ่ขยายจาก Unidimensional Rash Model โดยอาศัยการเชื่อมโยงคะแนนจากการศึกษา 4 มิติ ทำการศึกษาจากข้อมูลเชิงประจักษ์จากจำนวนนักเรียน 541 คน เพื่อประเมินผลสัมฤทธิ์ทางการเรียนวิทยาศาสตร์ โดยทำการเปรียบเทียบวิธีการศึกษา 3 วิธี คือ 1) Consecutive Unidimensional 2) Composite Unidimensional และ 3) Multidimensional ประมาณค่าพารามิเตอร์โดยใช้โปรแกรม ConQuest และตรวจสอบความสอดคล้องระหว่างโมเดลกับข้อมูลเชิงประจักษ์

ผลการวิจัยพบว่าวิธี Multidimensional มีความสอดคล้องกับข้อมูลเชิงประจักษ์มากกว่าวิธีอื่น และให้ค่าความเที่ยงในการประมาณค่าของผลสัมฤทธิ์ทางการเรียนวิทยาศาสตร์มากกว่าวิธีการ Consecutive Unidimensional และ Composite Unidimensional นอกจากพบว่าวิธีการแบบ Multidimensional มีความเหมาะสมกับการศึกษาผลสัมฤทธิ์ทางการเรียนวิทยาศาสตร์

Min (2003) ได้เปรียบเทียบและประเมินวิธีการเชื่อมโยงคะแนนตามโมเดล MIRT ที่ต่างกัน โดยพิจารณาความถูกต้องและความคงที่ของค่าการแปลงคะแนน (Transformations Metric) จากการศึกษาตามเงื่อนไข การกระจายความสามารถ (Ability Distribution) จำนวนของผู้สอบ และ

โครงสร้างของมิติ (Dimensional Structure) ที่ต่างกัน ซึ่งวิธีการที่นำมาเปรียบเทียบมี 3 วิธี คือ วิธี OLD (Oshima, Lee & Davy's Method) (ในบริบทของการศึกษา Test Characteristic Function; TCF) วิธี LL (Li & Lissize's Method) และวิธี M (Min's Method) ซึ่ง Min ได้พัฒนาวิธีการเชื่อมโยงคะแนนขึ้นเพื่อแก้ปัญหาค่าพารามิเตอร์การเลื่อนแกนที่เป็นสเกลลา (Scalar Dilation Parameter) โดยการแทนด้วยเมตริกซ์ (Diagonal Dilation Matrix) เพื่อให้สามารถเกิดการย่อส่วน (Compress) หรือขยาย (Dilate) สเกลให้อยู่ในตำแหน่งที่เหมาะสมมากขึ้นสำหรับการศึกษามิติที่มีความหลากหลาย ไม่ใช่การกำหนดค่าเพียงค่าใดค่าหนึ่งเท่านั้น โดยวิธี LL และ M มีการหมุนแกนแบบตั้งฉาก ในขณะที่วิธี TCF มีการหมุนแกนแบบไม่ตั้งฉาก เกณฑ์ที่ใช้ในการประเมินพิจารณาจาก BIAS และ RMSE

ผลการวิจัยพบว่า วิธีการเชื่อมโยงคะแนนขึ้นอยู่กับเงื่อนไขที่ศึกษา โดยเมื่อเทียบกับวิธี M พบว่า วิธี OLD มีความคลาดเคลื่อนของค่าอำนาจจำแนกน้อยกว่าสำหรับโครงสร้างที่ไม่ซับซ้อน (APSS) แต่มีค่าความคลาดเคลื่อนมากกว่าสำหรับโครงสร้างที่ซับซ้อนมากขึ้น (MS) และวิธี LL มีความคลาดเคลื่อนของค่าอำนาจจำแนกมากกว่าเมื่อเปรียบเทียบกับทั้ง 2 วิธี

นอกจากนี้พบว่า วิธี M ซึ่งเน้นการให้ค่าเมตริกซ์ (Diagonal Matrix) มีความคงที่ของการแปลงคะแนนสำหรับค่าพารามิเตอร์ที่มีความสัมพันธ์กับค่าความยากของข้อสอบ (Difficulty Related Parameters; MDIF) มากกว่าวิธี OLD เมื่อทำการสรุปผลโดยรวม พบว่า วิธี M ลดความคลาดเคลื่อนของค่าพารามิเตอร์อำนาจจำแนกมากกว่าอีก 2 วิธี อย่างไรก็ตาม จุดอ่อน ของวิธี M คือ ไม่สามารถคำนวณได้เมื่อศึกษาในมิติที่มีจำนวนมากเกินไป เนื่องจากปฏิสัมพันธ์ระหว่างคุณลักษณะเด่นของมิติข้ามชุดของแบบสอบ โดยเฉพาะคุณลักษณะของการหมุนแกนแบบตั้งฉาก ดังนั้นจำเป็นต้องมีการหมุนแกนของการเปลี่ยนแปลงมิติ จากการเปรียบเทียบแต่ละคู่ในแบบสอบปรับเทียบคะแนนไปยังแบบสอบฐาน

Reckase & Martineau (2004) ได้ศึกษาวิธีการสร้างสเกลแนวตั้งแบบพหุมิติตามโมเดล MIRT ชนิดโลจิสติก 3 พารามิเตอร์ เพื่อประยุกต์ใช้ศึกษาพัฒนาการภาคตัดขวางของผลสัมฤทธิ์ทางการเรียนวิทยาศาสตร์ของนักเรียน 4 ระดับชั้น ประกอบด้วยระดับชั้น 3-8 นั่นคือมีการออกแบบการเก็บรวบรวมข้อมูลรูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้แบบสอบร่วม โดยใช้ข้อสอบร่วมภายในระดับละ 25 ข้อ ซึ่งเนื้อหาทางวิทยาศาสตร์ที่จะมีความซับซ้อนประกอบด้วยการศึกษาทั้งหมด 15 มิติ ทำการ Calibrated ข้อมูลในแต่ละระดับชั้นโดยใช้โปรแกรม NOHARM (Fraser, 1988) เชื่อมโยงคะแนนในแต่ละระดับชั้นโดยใช้กระบวนการที่พัฒนาโดย Li & Lissitz (2000) & Min (2003) มาดัดแปลงสำหรับการศึกษาในครั้งนี้ โดยใช้เทคนิค 3 วิธี คือ 1) Orthogonal Procrustes Rotation 2) Translation Tranformation และ 3) Single Dilation เพื่อใช้ในการประมาณค่า 3 ค่า คือ 1) Non-Othogonal Procruster Rotation Matrix (T) ซึ่งได้พัฒนามาจาก Mulaik (1972) 2) Diagonal Dilation Matrix (K) โดยพัฒนาสมการมาจาก Min (2003) และ 3) Translation Vector (m) ซึ่งทำการพัฒนาสมการต่อจาก Lissitz (2000) เพื่อประยุกต์ใช้กับการศึกษาที่มีจำนวนมิติที่หลากหลาย นอกจากนี้มีการวิเคราะห์การจัดกลุ่มแบบลดหลั่น โดยใช้ Ward Method ในการระบุกลุ่มของข้อสอบที่มี

ลักษณะคล้ายคลึงกันให้อยู่ในกลุ่มเดียวกัน เพื่อพิจารณากลุ่มของข้อสอบที่สามารถจำแนกผลสัมฤทธิ์ทางการเรียนของนักเรียนได้ดีที่สุด

ผลการศึกษาพบว่า พัฒนาการของผลสัมฤทธิ์ทางการเรียนวิทยาศาสตร์ที่พิจารณาจากทักษะการปฏิบัติ มีลักษณะพัฒนาการที่ไม่เป็นแบบแผน นั่นคือมีรูปแบบของพัฒนาการที่หลากหลาย โดยเฉพาะพัฒนาการของระดับ 6 ไปยังระดับ 7 ซึ่งแบบสอบสะท้อนให้เห็นได้ถึงความแตกต่างของทักษะและความรู้ของระดับความสามารถในแต่ละระดับชั้นที่แตกต่างกันได้อย่างเด่นชัด แสดงว่าโมเดลแบบพหุมิติเป็นสิ่งจำเป็นที่สะท้อนให้เห็นถึงความซับซ้อนของการศึกษาพัฒนาการภาคตัดขวางของผลสัมฤทธิ์ทางการเรียนได้เป็นอย่างดี แต่ผลการศึกษามีข้อจำกัดของจำนวนข้อสอบร่วมที่มีจำนวนน้อยเกินไป (25 ข้อ) ในการที่จะดูความสัมพันธ์ของข้อสอบที่มีจำนวนไม่มากในแต่ละระดับกับความสัมพันธ์ของเนื้อหา จำเป็นต้องมีการศึกษาจำนวนของข้อสอบร่วมเพื่อให้การศึกษาพัฒนาการมีความถูกต้องมากขึ้น ผลการวิจัยดังกล่าวเป็นวิธีการหนึ่งที่สามารถจัดการกับปัญหาของวิธี M ที่พัฒนาโดย Min (2003) ซึ่งเป็นวิธีที่ไม่เหมาะสำหรับการศึกษาในมิติที่ซับซ้อนมากเกินไป โดยมีการพัฒนาการแปลงคะแนนแบบไม่ตั้งฉาก (Non-Orthogonal Procrustes Transformation) (Mulaik, 1972 อ้างใน Reckase & Martineau, 2004) จะทำการหมุนแกนเป็นไปโดยอัตโนมัติในการแปลงคะแนนในแต่ละมิติของเมตริกซ์ปรับเทียบคะแนนกับมิติของเมตริกซ์ฐาน ขึ้นอยู่กับความสอดคล้องที่สุดของมิติเมตริกซ์ปรับเทียบคะแนนกับมิติเมตริกซ์ฐาน ขณะเดียวกันการแปลงคะแนนแบบไม่ตั้งฉาก ช่วยลดความยุ่งยากไปในการคำนวณ โดยใช้การแปลงคะแนนแบบมุมแหลม (Oblique Transformation) อย่างไรก็ตามข้อจำกัดของการวิจัยนี้คือเหมาะสำหรับการศึกษาที่มีจำนวนมิติเพียง 2-3 มิติ ดังนั้นจำเป็นที่จะต้องมีการพัฒนาวิธีการปรับเทียบคะแนนในกรณีที่มีจำนวนมิติมากขึ้นพร้อมตรวจประสิทธิภาพของการศึกษาดังกล่าว

Reckase & Martineau (2006) ได้ศึกษาพัฒนาการภาคตัดขวางโดยทำการสร้างสเกลแนวตั้ง (Vertical Scaling) ตามโมเดล MIRT เน้นไปที่ความถูกต้องของมิติที่ศึกษาโดยอธิบายถึงกระบวนการตรวจสอบมิติของแบบสอบเพื่อแก้ปัญหาเกี่ยวกับกระบวนการอื่นที่เคยศึกษามา เพื่อบอกได้ว่าความสัมพันธ์ระหว่างผลการตอบข้อสอบได้วัดโครงสร้างเดียวกัน โดยศึกษาทั้งการจำลองข้อมูลและการใช้ข้อมูลจริง ทำการตรวจสอบความถูกต้อง 2 วิธี 1) ศึกษาจากเวคเตอร์ ตามแนวคิดของ Rackase et al. (2000) เพื่อค้นหามิติแฝงที่แท้จริง เมื่อองค์ประกอบที่แท้จริงอาจจะมีความสัมพันธ์กันสูงมากหรือไม่มีความสัมพันธ์กัน 2) ศึกษาจาก Scree Plot ที่วิเคราะห์ด้วย EFA ซึ่งเป็นแนวความคิดแบบดั้งเดิม เพื่อใช้กำหนดความเป็นมิติ

ผลการวิจัยจากการจำลองข้อมูลพบว่า เมื่อศึกษากับ 5 มิติ ซึ่งในแต่ละมิติไม่มีความสัมพันธ์กัน โดยศึกษาจาก 20 ตัวแปร และ 60 ตัวแปร พบว่าทั้งวิธีการศึกษาจาก เวคเตอร์ และการศึกษาจาก Scree Plot มีความถูกต้องในการประมาณค่าชุดของข้อมูล นอกจากนี้พบว่า การศึกษาจากเวคเตอร์ เหมาะสำหรับตรวจสอบมิติที่มีจำนวนมิติมากและจำนวนตัวแปรมากทั้งองค์ประกอบที่มีความสัมพันธ์กันและองค์ประกอบที่ไม่มีความสัมพันธ์กัน ตรงข้ามกับการศึกษาด้วย Scree Plot ที่เมื่อศึกษากับมิติเพิ่มมากขึ้นจะทำให้ความแม่นยำลดลง ส่วนผลการศึกษาจากข้อมูลจริง มีข้อจำกัดเพียง 3 มิติ จากการวิเคราะห์ EFA ด้วยการหมุนแกนแบบ Varimax และ Promax Rotation มี

ความเหมาะสม ส่วนศึกษาด้วยเวคเตอร์ ตัวแปรที่มี 6 ตัวแปร สามารถแปลงให้อยู่ในรูป 3 มิติ เพื่อให้ง่ายต่อการตีความหมาย

Yon (2006) ได้ศึกษาวิธีการเชื่อมโยงคะแนน 2 วิธี คือ วิธี TCF (Test Characteristic Function) และ วิธี NOP (Non-Orthogonal Procrustes) โดยศึกษาจากการจำลองข้อมูลและการใช้ข้อมูลจริงเพื่อประยุกต์ใช้ในการสร้างสเกลแนวตั้งแบบพหุมิติ (Multidimensional Vertical Scaling) ในการศึกษาพัฒนาการในระดับชั้นที่ต่ำกว่ากับระดับชั้นที่สูงกว่าทั้งหมด 3 ระดับชั้น กับเงื่อนไขที่ใช้ในการศึกษาคือ โครงสร้างมิติความสามารถ 2 ลักษณะ คือ APSS และ MS เกณฑ์ที่ใช้ในการพิจารณาความคลาดเคลื่อนของการเชื่อมโยงคะแนน คือ พิจารณาค่า BIAS RMSE และค่าความสัมพันธ์ของค่าพารามิเตอร์จากแบบสอบฐานและค่าพารามิเตอร์ที่ได้จากการแปลงคะแนนจากแบบสอบปรับเทียบคะแนน

ผลการวิจัย พบว่า วิธีการทั้งสอง มีแนวโน้มที่จะประมาณค่าที่มากกว่าความเป็นจริง แต่ให้ค่าความคลาดเคลื่อนของค่าพารามิเตอร์อำนาจจำแนกน้อยทั้งสองวิธี เมื่อพิจารณาในบริบทของความสัมพันธ์ และ BIAS วิธี NOP มีความเหมาะสมกับข้อมูลที่มีโครงสร้างที่ซับซ้อนมากขึ้น (MS) มากกว่าโครงสร้างที่ไม่ซับซ้อน (APSS) โดยพบว่าโครงสร้างมิติความสามารถของข้อมูลมีส่วนสำคัญในการประมาณค่าพารามิเตอร์ความยาก โดยการประมาณค่าพารามิเตอร์ความยากจากโครงสร้างที่ไม่ซับซ้อน มีความคลาดเคลื่อนน้อยอย่างมีนัยสำคัญทางสถิติมากกว่าข้อมูลที่มีโครงสร้างที่ซับซ้อน อย่างไรก็ตามแม้วิธีการ NOP จะให้ค่าความคลาดเคลื่อนสูงในการประมาณค่าพารามิเตอร์ความยากแต่ผลที่ได้จากการประมาณค่าอำนาจจำแนกและค่าความยากมีค่า RMSE ที่น้อย แต่ค่า RMSE มีผลอย่างไม่มีนัยสำคัญทางสถิติ เมื่อพิจารณาค่าความสัมพันธ์พบว่าค่าพารามิเตอร์ความยากของข้อสอบสามารถประมาณค่าได้ถูกต้องสำหรับข้อมูลที่มีโครงสร้างที่ไม่ซับซ้อน อย่างไรก็ตามโครงสร้างที่ซับซ้อนมากขึ้นเหมาะสมสำหรับวิธี NOP

Yon (2007) ได้ศึกษาเพื่อมุ่งถึงปัญหาการคำนวณความถูกต้องของเมตริกซ์ในการหมุนแกนเมื่อเกิดการเปลี่ยนแปลงโครงสร้าง ที่กำหนดให้แบบสอบมีจุดเน้นต่างกัน โดยเน้นที่การพิจารณาความแตกต่างระหว่างแบบสอบฐานกับแบบสอบปรับเทียบคะแนน เพื่อนำมาใช้ในการประมาณค่าพารามิเตอร์จากการจำลองข้อมูลใน 2 มิติ 2 ระดับชั้นที่ทำการศึกษา จากสถานการณ์การจำลองข้อมูลที่โครงสร้างมีการเปลี่ยนแปลง โดยพิจารณาจากค่า BIAS RMSE และค่าความสัมพันธ์ของค่าพารามิเตอร์จากแบบสอบฐานและค่าพารามิเตอร์ที่ได้จากการแปลงคะแนนจากแบบสอบปรับเทียบคะแนน

ผลการวิจัยพบว่า ค่าความสัมพันธ์และค่า RMSE มีค่าเพิ่มขึ้นเมื่อเปรียบเทียบจากแบบสอบฐานและการแปลงคะแนนจากแบบสอบปรับเทียบคะแนน แสดงให้เห็นว่าเมื่อโครงสร้างเกิดการเปลี่ยนแปลงจะส่งผลต่อความถูกต้องต่อการหมุนแกนโดยวิธี NOP มีความถูกต้องสำหรับโครงสร้างที่เกิดการเปลี่ยนแปลงได้ดี (Construct Shift)

จากการศึกษางานวิจัยที่เกี่ยวข้องในการนำทฤษฎี MIRT มาประยุกต์ใช้ในการปรับเทียบคะแนนแบบพหุมิติสามารถสรุปได้ดังตารางที่ 3

ตารางที่ 2.6 สรุปงานวิจัยที่เกี่ยวข้องกับการเชื่อมโยงคะแนนตามโมเดล MIRT

ผู้วิจัย	ประเด็นที่ศึกษา	เทคนิควิธี MIRT Linking	การออกแบบการวิจัย/ โปรแกรมการวิเคราะห์	เกณฑ์ที่ใช้ประเมิน	ประเด็นที่สามารถพัฒนาต่อได้
1. Hirsch (1989)	วิธีการปรับเทียบตามโมเดล MIRT ชนิด M2PLM จากการจำลองข้อมูลและการใช้ข้อมูลจริงสำหรับการวัดคุณลักษณะภายใน 2 มิติ	Rotation Indeterminacy (Orthogonal Procrustes Rotation)	-Multidimensional Horizontal Equating -Common - Examinee Design -โปรแกรม MIRTE	-Mean Difference -Mean Absolute Difference	1. พัฒนาเทคนิคในการประมาณค่าพารามิเตอร์เพื่อให้ผลการประเมินมีความคงที่ และอยู่ในระดับที่น่าพอใจ 2. ควรมีการศึกษาค่าพารามิเตอร์ Dilation
2. Oshima & Davey (1994)	ประเมินประสิทธิภาพการเชื่อมโยงคะแนนแบบ Compensatory MIRT Model จากการจำลองข้อมูลเพื่อศึกษาใน 2 มิติ	-Rotation Indeterminacy (Non-Orthogonal Procrustes Rotation) -พิจารณา ความแตกต่างของเวคเตอร์ค่าเฉลี่ยและความแตกต่างของเมตริกซ์ความแปรปรวนความแปรปรวนร่วม (V-C Matrix)	-Multidimensional Horizontal Equating -Common Items Design -โปรแกรม NOHARM	-RMSE	1. ควรมีการตรวจสอบผลที่ได้จากการใช้โปรแกรม NOHARM ให้เหมาะสมและสอดคล้องกับโมเดลมากยิ่งขึ้น 2. ควรที่จะมุ่งศึกษา Scaling Equation ที่แตกต่างกันหลายสมการ
3. Thompson et al. (1997)	พัฒนาวิธีการเชื่อมโยงคะแนนแบบ MIRT โดยการเชื่อมโยงคะแนนที่ไม่ต้องใช้ผู้สอบร่วมหรือข้อสอบร่วมภายใต้สถานการณ์การสอบที่มีแบบสอบแตกต่างกันโดยใช้การสุ่มกลุ่มผู้สอบที่มีขนาดใหญ่	-Rotation Indeterminacy	- Multidimensional Horizontal Equating -Equivalent Random Group Design	-BIAS -RMSE	1. พัฒนาวิธีการที่ใช้ระบุการจัดกลุ่มมวลเนื้อหาที่คล้ายคลึงกันจากแบบสอบที่แตกต่างกันหลายฉบับ 2. พัฒนาวิธีการ ให้มีความชัดเจนในเรื่องการออกแบบการเก็บรวบรวมข้อมูลกับข้อมูลจริง

ผู้วิจัย	ประเด็นที่ศึกษา	เทคนิควิธี MIRT Linking	การออกแบบการวิจัย/ โปรแกรมการวิเคราะห์	เกณฑ์ที่ใช้ประเมิน	ประเด็นที่สามารถพัฒนาต่อได้
4. Oshima, Davey & Lee (1997)	ศึกษาวิธีการเชื่อมโยงคะแนนแบบพหุมิติ โดยมุ่งศึกษาใน 2 มิติ การประมาณค่าความสามารถจะอาศัยเมตริกซ์การหมุนแกนและเวคเตอร์การแปลงคะแนน โดยใช้ฟังก์ชันทางคณิตศาสตร์ให้น้อยที่สุด	Non-Orthogonal Transformation Matrix	-Multidimensional Horizontal Equating -Common Items Design	-BIAS -RMSE	ควรพัฒนาวิธีการเชื่อมโยงคะแนนให้มีความสลับซ้อนน้อยที่สุดโดยอาศัยวิธีการเวคเตอร์ (Vector Approach)
5. Li & Lissitz (1997, 2000)	ประเมินประสิทธิภาพของเทคนิควิธีที่ใช้ในการเชื่อมโยงคะแนนแบบ MIRT Equating ชนิด M2PLM โดยการจำลองข้อมูล ให้เหมาะกับข้อมูลที่มีลักษณะแบบพหุมิติ ซึ่งมีการแจกแจงแบบปกติพหุ (MVN(0,1)) การจำลองข้อมูลในครั้งนี้ค่าพารามิเตอร์ของข้อสอบได้มาจากแบบสอบ ACT Form 24B จากการศึกษาของ Reckase ในปี 1985 เพื่อตรวจสอบความคลาดเคลื่อนของการประมาณค่าพารามิเตอร์	1) Orthogonal Procrustes Rotation 2) Translation Tranformation 3) Single Dilation **ตัวประมาณค่า** 1) ประมาณค่า k โดยวิธี Ratio of Eigenvalues, Trace, Least Square และ MTRs 2) ประมาณค่าพารามิเตอร์ในมิติของ m_1 และ m_2 โดยวิธี Least Square และ MTRs	-Multidimensional Horizontal Equating -Multidimensional Vertical Equating Anchor-Test Nonequivalent Design) โดยใช้จำนวนของข้อสอบร่วม 15 ข้อ และ 25 ข้อ -โปรแกรม TESTFACT ที่พัฒนาโดย Wilson et al. (1991)	-BIAS -RMSE	ควรมีการพัฒนาวิธีการเชื่อมโยงคะแนนในกรณีที่มีจำนวนมิติมากขึ้น พร้อมตรวจประสิทธิภาพของการศึกษาในเงื่อนไขที่สำคัญ เช่น 1) โมเดลการเชื่อมโยงคะแนนรูปแบบการเชื่อมโยงคะแนน 2) เทคนิคการเชื่อมโยงคะแนน 3) ตัวประมาณค่าพารามิเตอร์ 4) จำนวนมิติแฝง 5) จำนวนข้อสอบร่วม 6) ขนาดของกลุ่มตัวอย่าง เป็นต้น

ผู้วิจัย	ประเด็นที่ศึกษา	เทคนิควิธี MIRT Linking	การออกแบบการวิจัย/ โปรแกรมการวิเคราะห์	เกณฑ์ที่ใช้ประเมิน	ประเด็นที่สามารถพัฒนาต่อได้
6. Oshima, Davey & Lee (2000)	ทำการศึกษาวิธีการเชื่อมโยงคะแนนแบบ MIRT 4 วิธี โดยทำการประเมินประสิทธิภาพการเชื่อมโยงคะแนนโดยอาศัยข้อมูลจากการจำลองข้อมูล และค่าพารามิเตอร์ของข้อสอบได้มาจากการใช้ข้อมูลจริงเป็นแบบสอบที่มุ่งวัด 2 มิติ มี	1) Direct Method 2) Equated Function Method 3) Test Characteristic Function 4) Item Characteristic Function	-Multidimensional Horizontal Equating -Common Items Design โดยใช้ความยาวข้อสอบ 40 ข้อ -โปรแกรม NOHRAM	-BIAS -RMSE	1. ควรศึกษาเงื่อนไขอื่นที่เกี่ยวข้องกับการเชื่อมโยงคะแนน เช่น การศึกษาจำนวนมิติให้มากขึ้น 2. พัฒนาเทคนิควิธีในการเชื่อมโยงคะแนนแบบพหุมิติประกอบกับการตรวจสอบประสิทธิภาพจากเงื่อนไขดังกล่าว
7. Briggs (2003)	ทำการศึกษาวิธีการวัดแบบพหุมิติ โดยใช้โมเดล MRCM ที่แผ่ขยายมาจาก Unidimensional Rash Model โดยอาศัยการเชื่อมโยงคะแนนจาก 4 มิติ	1) Consecutive Unidimensional 2) Composite Unidimensional 3)Multidimensional	-Multidimensional Horizontal Equating -โปรแกรม ConQuest	เปรียบเทียบจากการพิจารณาค่า Deviance (G^2) Value	ควรมีการพัฒนาวิธีการปรับเทียบแบบพหุมิติที่มีการตรวจให้คะแนนมากกว่า 2 ค่า เพื่อเป็นประโยชน์ต่อการนำไปประยุกต์ใช้ในทักษะการปฏิบัติงาน
8. Min (2003)	พัฒนาวิธีการเชื่อมโยงคะแนน MIRT เพื่อแก้ปัญหาค่า “Scalar Dilation Parameter” โดยการแทนด้วย “Diagonal Dilation Matrix” กับการศึกษาพัฒนาการภาคตัดขวางทั้งหมด 3 มิติ	Non-Orthogonal 1) Orthogonal Procrustes Rotation 2) Transformation 3) Central Dilation	- Multidimensional Vertical Equating -Common Items Design -โปรแกรม TESTFACT ที่พัฒนาโดย Wilson et al. (1991)	-BIAS -RMSE	ควรมีการพัฒนาวิธีการเชื่อมโยงคะแนนในกรณีที่มีจำนวนมิติมากขึ้น พร้อมตรวจประสิทธิภาพของการศึกษา

ผู้วิจัย	ประเด็นที่ศึกษา	เทคนิควิธี MIRT Linking	การออกแบบการวิจัย/ โปรแกรมการวิเคราะห์	เกณฑ์ที่ใช้ประเมิน	ประเด็นที่สามารถพัฒนาต่อได้
9. Reckase & Martineau (2004)	ทำการศึกษาวิธีการปรับเทียบแนวตั้งแบบพหุมิติตามวิธีทฤษฎีตอบสนองข้อสอบชนิด M3PLM (MIRT Vertical Scaling) เพื่อนำมาประยุกต์ใช้กับการศึกษาพัฒนาการภาคตัดขวาง ทั้งหมด 15 มิติ	1) Non-Orthogonal Procrustes Rotation 2) Translation Tranformation 3) Single Dilation	- Multidimensional Vertical Equating -Anchor-Test Nonequivalent Group Design โดยใช้ความยาวแบบสอบ 50 ข้อ -โปรแกรม NOHARM (Fraser, 1988)	-BIAS -RMSE	1. ควรศึกษาจำนวนข้อสอบร่วมเท่าใดจึงเพียงพอต่อการเชื่อมโยงคะแนน 2. ควรศึกษาวิธีการรายงานผลระหว่าง “Single Scores” และ “Composite Score”
10. Reckase & Martineau (2006)	ศึกษาพัฒนาการภาคตัดขวางแบบ Multidimensional Vertical Scaling ตามแนวคิดของ MIRT ชนิด M2PLM ซึ่งเน้นไปที่ความถูกต้องของมิติที่ศึกษา โดยได้อธิบายถึงกระบวนการในการตรวจสอบมิติของแบบสอบโดยศึกษาทั้งการจำลองข้อมูลและการใช้ข้อมูลจริง	1. ศึกษาจากเวคเตอร์ (Vector Approach) จากแนวคิดของ Rackase et al. ในปี 2000 เพื่อค้นหามิติแฝงที่แท้จริง เมื่อองค์ประกอบที่แท้จริงอาจจะมีความสัมพันธ์กันสูงมากหรือไม่มีความสัมพันธ์กัน 2) ศึกษาจาก Scree Plot ซึ่งวิเคราะห์ด้วย EFA ซึ่งเป็นแนวความคิดแบบดั้งเดิม เพื่อใช้กำหนดความเป็นมิติ	- Multidimensional Vertical Equating -Anchor-Test Nonequivalent Group Design -โปรแกรม NOHARM (Fraser, 1988) และโปรแกรมในการวิเคราะห์ Factor Analysis	-BIAS -RMSE	1. ควรมีการศึกษาเปรียบเทียบประสิทธิภาพของวิธีการปรับเทียบแบบ Multidimensional Vertical Equating ในหลายๆ เงื่อนไข โดยเฉพาะจำนวนมิติที่ศึกษา 2 ควรพัฒนาวิธีการตรวจสอบมิติแฝงที่ทำการศึกษาให้มีความชัดเจนมากยิ่งขึ้น

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ผู้วิจัย	ประเด็นที่ศึกษา	เทคนิควิธี MIRT Linking	การออกแบบการวิจัย/ โปรแกรมการวิเคราะห์	เกณฑ์ที่ใช้ประเมิน	ประเด็นที่สามารถพัฒนาต่อได้
11. Yon (2006)	เปรียบเทียบวิธีการเชื่อมโยงคะแนนที่มีลักษณะการหมุนแกนแบบไม่ตั้งฉากเหมือนกัน 2 วิธี (TCF และ NOP) จากการจำลองข้อมูลและการใช้ข้อมูลจริงเพื่อนำไปประยุกต์ใช้ในการสร้างสเกลแนวตั้งแบบพหุมิติ (Multidimensional Vertical Scaling)	Non-Orthogonal Procrustes Rotation สำหรับวิธี NOP และ TCF	- Multidimensional Vertical Scaling -Anchor-Test Nonequivalent Group Design โดยใช้ข้อสอบร่วม 20 ข้อ -โปรแกรม GENDAT 5, BILOG-MG และ TESTFACT	BIAS, RMSE และค่าความสัมพันธ์ของค่าพารามิเตอร์จากแบบสอบฐานและค่าพารามิเตอร์ที่ได้จากการแปลงคะแนนจากแบบสอบปรับเทียบคะแนน	ควรมีการเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนนในกรณีที่มีลักษณะการหมุนแกนแตกต่างกัน เช่น วิธี NOP กับวิธี M โดยเน้นไปที่โครงสร้างมิติความสามารถที่แตกต่างกัน
12. Yon (2007)	มุ่งศึกษาถึงปัญหาการคำนวณความถูกต้องของเมตริกซ์ในการหมุนแกนเมื่อเกิดการเปลี่ยนแปลงโครงสร้าง (Construct Shift) โดยกำหนดให้แบบสอบมีจุดเน้นที่แตกต่างกัน โดยเน้นที่การพิจารณาความแตกต่างระหว่างแบบสอบฐานกับแบบสอบปรับเทียบคะแนน	Non-Orthogonal Procrustes Rotation	- Multidimensional Vertical Scaling -Anchor-Test Nonequivalent Group Design โดยใช้แบบสอบร่วม 20 ข้อ -โปรแกรม GENDAT 5, BILOG-MG และ TESTFACT	RMSE และค่าความสัมพันธ์ของค่าพารามิเตอร์จากแบบสอบฐานและค่าพารามิเตอร์ที่ได้จากการแปลงคะแนนจากแบบสอบปรับเทียบคะแนน	ควรมีการเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนนในกรณีที่มีลักษณะการหมุนแกนแตกต่างกัน โดยเฉพาะในเงื่อนไขของโครงสร้างมิติความสามารถทั้งแบบ APSS และ MS เนื่องจากงานวิจัยครั้งนี้ศึกษาเฉพาะโครงสร้างแบบ APSS เท่านั้น

กล่าวโดยสรุป จะเห็นว่าการนำโมเดล MIRT มาใช้ในการเชื่อมโยงคะแนน มีการพัฒนาอย่างต่อเนื่องตั้งแต่อดีตจนถึงปัจจุบัน ซึ่งการศึกษาในช่วงแรก (1989-1995) เป็นการศึกษาใน 2 มิติ และให้ผลการปรับเทียบที่ยังไม่น่าพอใจและวิธีการค่อนข้างมีความซับซ้อน ถือได้ว่าเป็นยุคบุกเบิกของการเชื่อมโยงคะแนนตามโมเดล MIRT ต่อมาเริ่มมีการศึกษามากขึ้นจนกระทั่งถึงปัจจุบันที่นำวิธีการตามโมเดล MIRT มาประยุกต์ใช้เพื่อให้เห็นผลในทางปฏิบัติในการศึกษาพัฒนาการ ซึ่งใช้วิธีการสร้างสเกลแนวตั้งแบบพหุมิติ (Multidimensional Vertical Equating) ซึ่งเริ่มมีการศึกษาในจำนวนมิติมากกว่า 2 มิติ เพื่อให้สอดคล้องกับสถานการณ์การวัดในปัจจุบันที่ต้องการวัดคุณลักษณะแฝงที่มีหลายมิติ นอกจากนี้การเชื่อมโยงคะแนนตามโมเดล MIRT มีการศึกษาเฉพาะในต่างประเทศ ส่วนในประเทศไทยถือเป็นเนื่องใหม่ที่ไม่มีผู้ศึกษาโดยตรง ดังนั้นจึงเป็นสิ่งจำเป็นที่ต้องศึกษางานวิจัยที่เกี่ยวข้องกับการนำโมเดล MIRT มาประยุกต์ใช้ในการเชื่อมโยงคะแนน เพื่อเป็นสารสนเทศสำคัญในการพัฒนาองค์ความรู้ดังกล่าวในประเทศ

ดังนั้นผู้วิจัยจึงสนใจศึกษาในประเด็นของการเชื่อมโยงคะแนนตามโมเดล MIRT ไปประยุกต์ใช้ในการศึกษาพัฒนาการความสามารถของผู้เรียนในรูปแบบของการสร้างสเกลแนวตั้งแบบพหุมิติ (Multidimensional Vertical Scaling) โดยมีแนวคิดสำคัญ วิธีการนำไปประยุกต์ใช้ และงานวิจัยที่เกี่ยวข้อง ดังรายละเอียดที่จะได้นำเสนอต่อไป

ตอนที่ 3 การสร้างสเกลแนวตั้งแบบพหุมิติ (Multidimensional Vertical Scaling)

การเสนอวิธีการนำโมเดล MIRT ไปประยุกต์ใช้เพื่อศึกษาพัฒนาการความสามารถของผู้เรียนในรูปแบบของการสร้างสเกลแนวตั้งแบบพหุมิติ (Multidimensional Vertical Scaling) ผู้วิจัยได้นำเสนอที่มาและความสำคัญ มโนทัศน์เบื้องต้นเกี่ยวกับการสร้างสเกลแนวตั้งแบบพหุมิติ ความหมายของพัฒนาการ กระบวนการในการเชื่อมโยงคะแนนตามการสร้างสเกลแนวตั้งแบบพหุมิติเพื่อนำไปประยุกต์ใช้ในการศึกษาพัฒนาการ ตลอดจนงานวิจัยที่เกี่ยวข้อง เพื่อนำไปสู่ประเด็นในการพัฒนาองค์ความรู้ให้สามารถนำไปปฏิบัติได้จริงต่อไป โดยมีรายละเอียดดังนี้

จากการศึกษางานวิจัยทั้งในประเทศและต่างประเทศ นิยมนำวิธีการปรับเทียบแนวตั้ง (Vertical Equating) ตามวิธีทฤษฎีการตอบสนองข้อสอบ แบบโลจิสติก 3 พารามิเตอร์ ในการศึกษาพัฒนาการหรือความก้าวหน้าในการเรียน ซึ่งใช้แบบสอบหลายระดับกับผู้สอบที่มีความสามารถต่างกัน เนื่องจากให้ผลการปรับเทียบที่มีความเพียงพอและน่าพอใจกว่าวิธีอื่น (Kolen, 1980; Kolen & Brennan,1995; พรพิมล นาคเวช, 2537; วรรณ์ดี แสงประทีปทอง, 2538; ภัทรพร เกษสังข์, 2546; พัชรี จันทร์เพ็ง, 2547) แต่ปัญหาที่มักจะเกิดขึ้นเสมอเมื่อปรับเทียบคะแนนคือมีข้อจำกัดในเนื้อหาที่ต้องการศึกษาพัฒนาการเพียงเรื่องใดเรื่องหนึ่งเท่านั้น ซึ่งไม่สอดคล้องกับสถานการณ์การสอบในปัจจุบัน เนื่องจากข้อสอบหลายฉบับที่นำมาปรับเทียบกันจะต้องมีลักษณะเนื้อหาเพียงมิติเดียว ซึ่งวัดคุณลักษณะเด่นเพียงคุณลักษณะเดียว จะเห็นได้ว่าโมเดลหลายๆ โมเดลที่ใช้ในการศึกษาความก้าวหน้าหรือพัฒนาการ

ของผู้เรียนถูกพัฒนาขึ้นภายใต้พื้นฐานของการปรับเทียบแนวตั้งแบบเอกมิติ (Vertical Equating) ซึ่งจากการศึกษางานวิจัยที่ผ่านมายังคงมุ่งแปรความหมายของพัฒนาการในลักษณะของความแปรผัน (Variation) อัตราเพิ่มขึ้น (Acceleration) อัตราลดลง (Deceleration) หรืออัตราเร็วก้าวกระโดด (Jump) เป็นต้น โดยในการวัดดังกล่าวยังเป็นประเด็นปัญหาที่ต้องมีการศึกษา ซึ่งในบริบทของการวัดการเปลี่ยนแปลงจากสเกลหนึ่งไปยังอีกสเกลหนึ่งให้อยู่ในสเกลเดียวกัน โดยเฉพาะการเชื่อมโยงคะแนนจากระดับชั้นหนึ่งไปยังอีกระดับชั้นหนึ่ง คำถามที่สำคัญคือการใช้เทคนิคการปรับเทียบแนวตั้งแบบเอกมิติมีความตรง (Validity) หรือไม่ ต่อการนำมาใช้ในการศึกษาพัฒนาการข้ามกลุ่มความสามารถที่ต่างกัน

ด้วยต้องการที่จะแก้ไขจุดอ่อนที่สำคัญ ภายใต้ข้อตกลงเบื้องต้นของความเป็นเอกมิติ ได้มีการพัฒนาวิธีการวัดพัฒนาการในลักษณะที่มีหลายมิติ โดยอาศัยแนวคิดของทฤษฎีการตอบสนองข้อสอบแบบพหุมิติเข้ามาช่วยแก้ปัญหาดังกล่าว (MIRT Vertical Scaling) เพื่อนำมาใช้ในการเชื่อมโยงคะแนนข้ามระดับชั้นโดยใช้แบบสอบที่มีความยากต่างกัน ซึ่งจากการศึกษาของ Li & Lissiz (2000) Min (2003) & Reckase & Martineau (2004,2006) พบว่ายังพบปัญหาในประเด็นของการนำไปใช้ในทางปฏิบัติ สำหรับการนำไปใช้ในการประเมิน **Large-Model** จึงจำเป็นต้องมีการพัฒนาการสร้างสเกลแนวตั้งแบบพหุมิติให้มีความถูกต้องและสามารถนำไปใช้ในการศึกษาในทางปฏิบัติมากยิ่งขึ้น

การศึกษาครั้งนี้ผู้วิจัยได้นำเสนอมโนทัศน์เบื้องต้นเกี่ยวกับการสร้างสเกลแนวตั้งแบบพหุมิติ (Multidimensional Vertical Scaling) พร้อมทั้งได้นำเสนองานวิจัยที่เกี่ยวข้อง โดยมีรายละเอียดดังนี้

3.1 มโนทัศน์เบื้องต้นเกี่ยวกับการสร้างสเกลแนวตั้งแบบพหุมิติ

การศึกษาในสถานการณ์การทดสอบในปัจจุบันพบว่ากำลังให้ความสำคัญกับการนำวิธีปรับเทียบสเกลแนวตั้ง (Vertical Scaling) ไปประยุกต์ใช้ทางด้านการศึกษาเพื่อวัดพัฒนาการของผู้เรียน ดังพระราชบัญญัติการศึกษาแห่งชาติ พ.ศ. 2542 และแก้ไขเพิ่มเติม (ฉบับที่ 2) พ.ศ. 2545 ที่มุ่งเน้นการประเมินความก้าวหน้าการเรียนรู้ของผู้เรียน โดยรายงานกล่าวแนะนำว่า การนำวิธีการสร้างสเกลแนวตั้งมาช่วยในการเชื่อมโยงคะแนนเพื่อศึกษาการเพิ่มขึ้นของทักษะการปฏิบัติ (Thum, 2003 อ้างใน Yon, 2006) ซึ่งสามารถนำสเกลแนวตั้ง (Vertical Scale) มาใช้ในการประเมินแนวโน้มของการเพิ่มขึ้นและศึกษาพัฒนาการนั่นเอง

การสร้างสเกลแนวตั้ง (Vertically Scaled) ถือได้ว่าเป็นมาตรฐานสำคัญที่บ่งชี้ถึงผลสัมฤทธิ์ทางการเรียนซึ่งถือเป็นหัวใจสำคัญของการออกแบบในโมเดลการวัดมูลค่าเพิ่มทางการศึกษา ส่งผลให้แบบสอบวัดผลสัมฤทธิ์ทางการเรียนเน้นในเรื่องการประเมินการเพิ่มขึ้นและเน้นการศึกษาพัฒนาการ โดย Thum (2003) กล่าวว่า โรงเรียนในเขตต่างๆ จะใช้โมเดลการเพิ่มขึ้นของพัฒนาการ โดยเน้นไปที่คะแนนที่ได้แปลงมาจากการปรับเทียบแนวตั้งของแบบสอบจากแบบสอบวัดผลสัมฤทธิ์ทางการเรียนปัจจุบันโดยส่วนใหญ่มีการใช้กระบวนการวัดโมเดลพัฒนาการ ซึ่งเป็นการวัดระดับอันตรภาคชั้นเพื่อศึกษาข้ามกลุ่มหรือข้ามระดับความสามารถ ความตรงของการสรุปอ้างอิงเกี่ยวกับพัฒนาการของผู้เรียน

มีความเป็นไปได้เมื่อพัฒนาการ ถูกพิจารณาในบริบทของการศึกษาทางด้านเนื้อหาในแต่ละระดับชั้น ซึ่งโมเดลทั้งหมดจะใช้คะแนนของนักเรียนจากระดับชั้นหนึ่งไปยังอีกระดับชั้นหนึ่งเพื่อวัดการเพิ่มขึ้น แต่อาจจะง่ายเกินไป เนื่องจากแบบสอบในระดับชั้นที่ต่างกันอาจจะวัดคุณลักษณะที่แตกต่างกัน (Reckase, 2004) เมื่อพิจารณาตามหลักคณิตศาสตร์ เช่น นักเรียนในระดับมัธยมศึกษาปีที่ 1 จะวัดทักษะทางด้านเรขาคณิต ขณะที่นักเรียนในระดับชั้นมัธยมศึกษาปีที่ 3 จะเน้น ไปที่ทักษะการแก้ปัญหา และทักษะทางด้านพีชคณิต เป็นต้น อย่างไรก็ตามผลที่ได้จะถูกรายงานเพื่อให้มีความหมายว่าวัดในสิ่งเดียวกัน ดังคำกล่าวของ Lord (1963 อ้างใน Yon, 2007) ว่า “การวัดการเปลี่ยนแปลง” มีความหมายว่า “เป็นการระบุคะแนนที่ได้จาก 2 สเกล โดยนำมาใช้ในการแปลความอย่างมีความหมาย…” นั่นคือแบบสอบที่ใช้นำมาเปรียบเทียบต้องวัดในคุณลักษณะเดียวกันบนสเกลที่สามารถเปรียบเทียบกันได้ ตามสมมติฐานภายใต้กระบวนการวัดแนวตั้งของแบบสอบวัดผลสัมฤทธิ์ทางการเรียน ซึ่งต้องอยู่ในเนื้อหาเดียวกันภายใต้การวัดที่หลากหลาย (Multiple Measurement) โดยต้องใช้วิธีการที่มีความถูกต้อง เป็นธรรมเพื่อเป็นตัวแทนของการวัดให้อยู่ในสเกลเดียวกัน ซึ่งคุณสมบัติดังกล่าวเรียกว่า “ความไม่แปรเปลี่ยนของการวัด (Measurement Invariance)”

สำหรับในการสร้างสเกลแนวตั้ง ซึ่งมีโครงสร้างจากแบบสอบที่มีลักษณะเฉพาะในแต่ละระดับการศึกษาที่มีความหลากหลาย การศึกษาความไม่แปรเปลี่ยนของการวัดจึงเป็นสิ่งจำเป็น นั่นคือ แบบสอบที่ใช้เป็นรายบุคคล (ต้องวัดในแต่ละลักษณะให้อยู่ในโครงสร้างเดียวกันแต่มีระดับที่ต่างกัน เช่น ผลการสอบวิทยาศาสตร์ของนักเรียนระดับมัธยมศึกษาปีที่ 4 เน้นผลสัมฤทธิ์ทางการเรียนด้านความรู้ความเข้าใจ ในขณะที่ระดับมัธยมศึกษาปีที่ 5 คะแนนที่ได้บ่งชี้ถึงผลสัมฤทธิ์ทางการเรียนด้านความรู้ความเข้าใจ ถ้าต้องการแทนคะแนนจากแบบสอบทั้งสองฉบับให้อยู่ในสเกลเดียวกัน และมีการแปลความหมายการเพิ่มขึ้นหรือพัฒนาการที่แตกต่างกันไปตามระดับชั้นที่ทำการศึกษา จะต้องตั้งสมมติฐานว่าความรู้และทักษะต้องวัดทั้ง 2 ระดับชั้นที่คล้ายกันอย่างเพียงพอเพื่อให้สอดคล้องกับคะแนนที่ได้อย่างมีความหมายเพื่อให้อยู่ในหน่วยการวัดที่มีลักษณะต่อเนื่องโดยสามารถเปรียบเทียบกันได้

ตามที่การศึกษาการปรับเทียบแนวตั้งให้ผลที่มีความถูกต้องนั้นจะต้องมีเนื้อหาคล้ายกันเพียงพอเมื่อศึกษาในระดับชั้นที่ต่างกัน นั่นคือ ทักษะการปฏิบัติของแบบสอบในระดับชั้นที่ต่ำกว่า (Lower-Grade Test) เป็นเนื้อหาที่เกี่ยวเนื่องกับระดับชั้นถัดไปหรือกล่าวอีกนัยหนึ่งคือ เป็นแบบสอบก่อนเรียน (Pre-Test) ของแบบสอบในระดับชั้นถัดไป ถ้าแบบสอบชุดเดียวเป็นทั้งแบบสอบหลังเรียนสำหรับระดับชั้นก่อนหน้านี้และแบบสอบก่อนเรียนสำหรับระดับชั้นถัดไป ดังนั้นความแตกต่างระหว่างระดับชั้นที่อยู่ต่ำกว่าและระดับชั้นที่สูงกว่าของสเกลการสร้างสเกลแนวตั้ง สามารถตีความหมายได้เหมือนกับการเพิ่มขึ้นของผลสัมฤทธิ์ทางการเรียนเมื่ออยู่ในระดับชั้นที่สูงขึ้น ภายใต้ข้อตกลงเบื้องต้นว่ามีความเหลื่อมล้ำของหลักสูตรเนื้อหาอย่างเพียงพอ

ในปี ค.ศ. 2004 Matineau ได้จำลองข้อมูลโดยตรวจสอบความแกร่งของความไม่แปรเปลี่ยนของการวัด ซึ่งการศึกษามีประโยชน์ในการตรวจสอบผลของการเปลี่ยนโครงสร้าง (Construct-Shift) จากปีหนึ่งไปยังปีถัดไป ซึ่งผลที่ได้บ่งชี้ว่า ในสถานการณ์จริงมีความเป็นไปได้ที่จะมีการเพิ่มคะแนนพัฒนาการและในขณะเดียวกันมีความเป็นไปได้ที่จะไม่มีการเพิ่มคะแนนพัฒนาการ กล่าวสั้นๆ คืออาจเกิดความผิดพลาดถ้ามีการเปลี่ยนแปลงเชิงโครงสร้างเกิดขึ้น ซึ่งสามารถใช้วิธีการสร้างสเกลแนวตั้งมาช่วยแก้ปัญหาดังกล่าวได้ ถ้าประเด็นดังที่กล่าวมาแล้วไม่มีความถูกต้องและเพียงพอ

สิ่งหนึ่งที่มีความเป็นไปได้ที่ช่วยแก้ปัญหาผลของการเปลี่ยนแปลงเชิงโครงสร้าง คือ การใช้โมเดลแบบพหุมิติ โดยใช้แบบสอบในระดับชั้นที่ต่างกัน ศึกษาในแต่ละคุณลักษณะหรือในแต่ละมิติซึ่งมีเนื้อหาร่วมกัน วิธีการวัดแบบพหุมิติควรที่จะให้สารสนเทศที่มากขึ้น อย่างไรก็ตามการสร้างสเกลแนวตั้งจะต้องใช้วิธีการที่มีความซับซ้อนมากขึ้น

3.1.1 การสร้างสเกลแนวตั้ง (Vertical Scaling)

การสร้างสเกลแนวตั้ง เป็นศัพท์ทั่วไปในกระบวนการของแบบสอบที่มีระดับความยากที่ต่างกันให้ไปอยู่ในสเกลเดียวกัน (Petersen, Kolen, & Hoover, 1989) โดยจุดมุ่งหมายหลัก คือ การสร้างสเกลเดียวกันแก่ผู้ใช้แบบสอบ เช่น โรงเรียนที่อยู่ในระดับเขตซึ่งมีความหมายที่ดีกว่าการศึกษาพัฒนาการของผลสัมฤทธิ์ทางการเรียนข้ามปีและระดับชั้น

การสร้างสเกลแนวตั้งเป็นการนำเสนอการปรับ (Calibrate) รูปแบบของแบบสอบ (เช่น ระดับ (Level)) เพื่อมุ่งใช้ในการเรียงลำดับ (Ordinal) ซึ่งได้ให้ความหมายของการเปลี่ยนแปลงคะแนนของนักเรียนในระดับชั้นใดระดับชั้นหนึ่งเพื่อไปสู่สเกลการวัดร่วมกัน สามารถใช้ในการประเมินพัฒนาการได้โดยตรงสำหรับผู้สอบเป็นรายบุคคล แบบสอบที่นำมาใช้ในการสร้างสเกลแนวตั้งมุ่งที่จะนำไปใช้อ้างอิงให้เกิดความตรง (Valid Inferences) เกี่ยวกับพัฒนาการ โดยการแผ่ขยายพิสัยความกว้างของเนื้อหา เพื่อใช้ศึกษาหลายระดับชั้น โดยเฉพาะบริบทของรายวิชา และมุ่งที่การประเมินพัฒนาการจากระดับชั้นหนึ่งไปยังอีกระดับชั้นหนึ่ง โดยมีสเกลที่จะสามารถนำมาเปรียบเทียบกันได้ กล่าวโดยสรุป คือ การแปลงการสร้างสเกลแนวตั้ง คือ การบ่งชี้ทักษะการปฏิบัติของผู้เรียนบนสเกลที่มีความต่อเนื่อง นั่นคือ การแผ่ขยายระนาบที่มีความหลากหลายของระดับของแบบสอบ (Petersen, Kolen, & Hoover, 1989)

วิธีการหนึ่งที่ได้รับความนิยมในการสร้างสเกลแนวตั้ง คือการใช้โมเดล IRT ตั้งอยู่บนพื้นฐานของการศึกษาการปรับเทียบคะแนนและทฤษฎีของ IRT ซึ่งผู้ดำเนินการสร้างข้อสอบจะสร้างสเกลแนวตั้งมีการข้ามสเกลจากระดับประถมศึกษาไปยังระดับมัธยมศึกษา คะแนนที่ได้จากการสอบจะมีระดับที่ต่างกันในการแปลงระดับความสามารถทั้งหมด สามารถรายงานความถูกต้องได้ เมื่ออยู่ในสเกลเดียวกัน เช่น ในสถานการณ์ของการแปลงคะแนนในระดับที่ 4 โดยใช้แบบสอบ SAT-9 (The Ninth Edition of the Stanford Achievement Test) ทำการรายงานให้อยู่ในสเกลเดียวกันเช่นเดียวกับนักเรียนในระดับที่ 7 ทำแบบสอบ SAT-9 เช่นเดียวกันกับคะแนนของนักเรียนในระดับที่ 7 ที่ทำแบบสอบในระดับ

ที่ 6 จะรายงานให้อยู่ในสเกลเดียวกันเช่นเดียวกับนักเรียนในระดับที่ 7 ที่ทำแบบสอบระดับ 7 ซึ่งสิ่งที่เป็นประโยชน์ของการจัดรูปแบบให้อยู่ในสเกลเดียวกัน คือ การแผ่ขยายระนาบระดับทั้งหมดของแบบสอบซึ่งตำแหน่งของผู้สอบจะอยู่ในสเกลโดยไม่ขึ้นกับระดับใดระดับหนึ่งของแบบสอบโดยเฉพาะ เมื่อใดก็ตามที่ผู้สอบทำแบบสอบในระดับชั้นที่ศึกษาหรือแบบสอบในระดับชั้นที่ต่ำกว่า/ระดับชั้นที่สูงกว่า คะแนนที่ได้จากสเกลต้องตั้งอยู่บนข้อตกลงเบื้องต้นว่าอยู่บนสเกลเดียวกัน และสามารถนำคะแนนดังกล่าวมาเปรียบเทียบกัน

การสร้างสเกลแนวตั้ง (Vertical Scaling) เป็นวิธีการสร้างสเกลกับการออกแบบที่มีความเฉพาะเจาะจงในการเก็บรวบรวมข้อมูล มีการออกแบบโดยใช้การร่วมทั้งหมด 3 รูปแบบ คือ 1) การใช้ข้อสอบร่วม (Common-Item) 2) การใช้กลุ่มที่มีความเท่าเทียมกัน (Random Equivalent Group) และ 3) การออกแบบโดยใช้การสร้างสเกลร่วมของแบบสอบ (Scaling-Test Design) (Kolen & Brennan, 2004) สำหรับการสร้างสเกลร่วมตามกระบวนการทาง IRT อยู่ภายใต้ข้อตกลงเบื้องต้นของความเป็นเอกมิติ ข้ามสเกลแนวตั้ง โดยมีวิธีการที่นิยมเช่น 1) วิธี Mean/Mean (Loyd & Hoover, 1980) 2) วิธี Mean/Sigma (Marco, 1977) 3) วิธี Test Characteristic Curves (Haebara, 1980) & (Stocking & Lord, 1983) และ 4) วิธี Logit (Ogasawara, 2001)

ในการพัฒนาสเกลแนวตั้งข้ามระดับชั้น เป็นไปได้ยากที่จะหลีกเลี่ยงความเป็นพหุมิติของแบบสอบเนื่องจากต้องมีความเฉพาะเจาะจงในแต่ละระดับชั้นในขณะเดียวกันเมื่อข้ามระดับชั้นจะต้องมีโครงสร้างที่ร่วมกัน (Construct Common) (Camilli, 1988, 1999; Schulz & Nicewander, 1997) ซึ่งนำไปสู่การนำวิธีการแบบพหุมิตินำมาใช้ในการสร้างสเกลแนวตั้งเพื่อให้สอดคล้องกับบริบทของสภาพความเป็นจริงดังที่จะได้กล่าวในรายละเอียดต่อไป

3.1.2 ความเป็นพหุมิติและการสร้างสเกลแนวตั้ง (Multidimensional and Vertical Scaling)

นักวิจัยส่วนใหญ่ยอมรับว่าข้อมูลที่ได้จากแบบสอบทางด้านการศึกษาและจิตวิทยาจะมีลักษณะเป็นแบบพหุมิติ (Ackerman, 1991; Traub, 1983) โดยจากการวิจัยแสดงให้เห็นว่าเมื่อทราบว่ามีลักษณะเป็นแบบพหุมิติ แต่มีการใช้โมเดลแบบเอกมิติ อาจทำให้เกิดความคลาดเคลื่อนในการวัดส่งผลให้การสรุปอ้างอิงไม่ถูกต้อง เกี่ยวกับความสามารถของนักเรียนในบางวิชาที่ประสบปัญหาดังกล่าว (Walker & Beretvas, 2000) ดังนั้นจึงเป็นสิ่งสำคัญในการศึกษากรณีที่มีคุณลักษณะแฝงหลายด้าน ซึ่งการใช้โมเดล MIRT มาช่วยเป็นวิธีวิทยาที่สำคัญที่เป็นประโยชน์ในการประเมินทางด้านการศึกษาและจิตวิทยา (Reckase, 1997)

จากการศึกษาที่ผ่านมามีการตรวจสอบผลของความเป็นพหุมิติภายใต้แนวคิดของ การปรับเทียบคะแนน (Equating) แต่มีงานวิจัยน้อยมากที่ศึกษาการสร้างสเกลแนวตั้ง (Vertical Scaling) โดยการศึกษาของ Brogan & Yen (1983) ได้เปรียบเทียบค่าพารามิเตอร์ของการสร้างสเกลแนวตั้ง (Vertical Scale) ระหว่างการจำลองข้อมูลในสถานการณ์ที่เป็นเอกมิติและพหุมิติ นอกจากนี้ยังมี

งานวิจัยที่ทำการสนับสนุนในเรื่องดังกล่าว (Camilli, 1988; Camilli, Yamamoto & Wang, 1993; Loyd & Plake, 1987; Patz, Yao, Chia, Lewis, & Hoskens, 2003)

นอกจากนี้มีการศึกษาความไม่แปรเปลี่ยนของการวัด สำหรับการสร้างสเกลแนวตั้ง โดย Skaggs & Lissitz (1988) ได้ศึกษาจากการจำลองข้อมูลภายในสถานการณ์ที่มีความเป็นเอกมิติ และสเกลของการวัดมีความแปรเปลี่ยนไปตามกลุ่มความสามารถของผู้เรียน แต่เมื่อมีการศึกษากับข้อมูลที่มีลักษณะเป็นแบบพหุมิติ พบว่าไม่มีความแปรเปลี่ยนไปตามกลุ่ม ดังนั้นผลการวิจัยดังกล่าวจึงสรุปว่าการสร้างสเกลแนวตั้ง (Vertical Scaling) เหมาะสำหรับข้อมูลที่มีลักษณะเป็นพหุมิติ

3.1.3 ความหมายของการสร้างสเกลแนวตั้ง (Multidimensional Vertical Scaling)

Hambleton & Swaminathan (1984), Kolen & Brennon (2004), ศิริชัย กาญจนวาสี (2545) ได้ให้ความหมายของการปรับเทียบแนวตั้ง (Vertical Equating) ไว้คล้ายคลึงกัน ซึ่งเมื่อนำมาพิจารณาในความหมายของการสร้างสเกลแนวตั้ง (Vertical Scaling) สามารถสามารถสรุปได้ดังนี้

การสร้างสเกลแนวตั้ง (Vertical Scaling) เป็นการปรับเทียบคะแนนระหว่างแบบสอบต่างฉบับกัน เมื่อแบบสอบต่างฉบับมุ่งวัดลักษณะเดียวกัน แต่มีระดับความยากแตกต่างกันและกลุ่มผู้สอบมีการแจกแจงความสามารถอยู่คนละประชากรหรือมีความสามารถแตกต่างกัน เหมาะกับสถานการณ์ที่มีความจำเป็นต้องสร้างแบบสอบต่างฉบับวัดเนื้อหาเดียวกัน มุ่งวัดความสามารถของผู้สอบที่ต่างระดับกัน เพื่อปรับเทียบว่าคะแนนที่สอบได้จากฉบับหนึ่งเทียบเป็นเท่าไรของฉบับอื่นที่วัดต่างระดับกัน นำไปใช้ในการปรับเทียบคะแนนระหว่างแบบสอบวิชาเดียวกันแต่ต่างระดับชั้น

เมื่อศึกษาความหมายของการปรับเทียบเสกลแนวตั้งแบบพหุมิติ (Multidimensional Vertical Scaling) ยังไม่มีผู้ให้ความหมายโดยตรง แต่จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง สามารถสรุปได้ว่า **การสร้างสเกลแนวตั้งแบบพหุมิติ (Multidimensional Vertical Scaling)** หมายถึง การปรับเทียบคะแนนระหว่างแบบสอบต่างฉบับกัน เมื่อแบบสอบต่างฉบับมุ่งวัดลักษณะเดียวกันแต่ศึกษาในหลายมิติ ซึ่งมีระดับความยากแตกต่างกันและกลุ่มผู้สอบมีการแจกแจงความสามารถอยู่คนละประชากรหรือมีความสามารถแตกต่างกัน เหมาะกับสถานการณ์ที่จำเป็นต้องสร้างแบบสอบต่างฉบับวัดเนื้อหาหลายมิติแฝงโดยมุ่งวัดความสามารถของผู้สอบที่ต่างระดับกัน เพื่อปรับเทียบว่าคะแนนที่สอบได้จากฉบับหนึ่งเทียบเป็นเท่าไรของฉบับอื่นที่วัดต่างระดับกัน โดยแยกศึกษาในแต่ละมิติ รวมถึงศึกษาเป็นคะแนนรวมที่ (Composite Scores) เช่น การปรับเทียบสเกลระหว่างแบบสอบวัดความถนัด ที่ประกอบด้วยการวัดทางด้านคณิตศาสตร์และการวัดทางด้านภาษา ซึ่งในกรณีนี้เป็นการศึกษาใน 2 มิติ ของระดับชั้น ม.1-ม.3 ในการปรับเทียบสเกลกันนั้น แบบสอบแต่ละฉบับจึงมีระดับความยากแตกต่างกันไป รวมทั้งกลุ่มผู้สอบแต่ละฉบับมีการแจกแจงความสามารถอยู่ต่างประชากร หรือมีการแจกแจงความสามารถที่อยู่ในระดับที่แตกต่างกัน

เมื่อทราบถึงความหมายของพัฒนาการในบริบทของการศึกษาภาคตัดขวางจากระดับที่แตกต่างกัน สิ่งจำเป็นที่จะทำการศึกษาพัฒนาการความสามารถของผู้เรียนในการนำวิธีการปรับเทียบตามโมเดล MIRT มาใช้ในการศึกษาภาคตัดขวาง คือ การออกแบบการเก็บรวบรวมข้อมูล ซึ่งใน

สถานการณ์โดยส่วนใหญ่จะมีลักษณะการออกแบบการเก็บรวบรวมข้อมูลโดยใช้รูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้แบบสอบร่วม และสิ่งที่จำเป็นอีกประการหนึ่งคือ กระบวนการปรับเทียบคะแนนแบบ MIRT ในบริบทของการสร้างสเกลแนวตั้ง (Multidimensional Vertical Scaling) ซึ่งแม้ว่าการเชื่อมโยงคะแนนดังกล่าวจะมีความซับซ้อนกว่าการเชื่อมโยงคะแนนในแนวนอนและการสร้างสเกลแนวตั้งแบบเอกมิติแต่รายละเอียดโดยส่วนใหญ่มีลักษณะที่คล้ายคลึงกันทั้งในแง่ทฤษฎีและการปฏิบัติ แต่จะต่างกันในแง่ของการออกแบบและการแปลความหมายของคะแนนที่ปรับเทียบได้ ในส่วนนี้ผู้วิจัยจะนำเสนอกระบวนการในการสร้างสเกลแนวตั้งแบบพหุมิติ (Multidimensional Vertical Scaling) เพื่อนำไปประยุกต์ใช้ในการศึกษาพัฒนาการ โดยมีรายละเอียดดังนี้

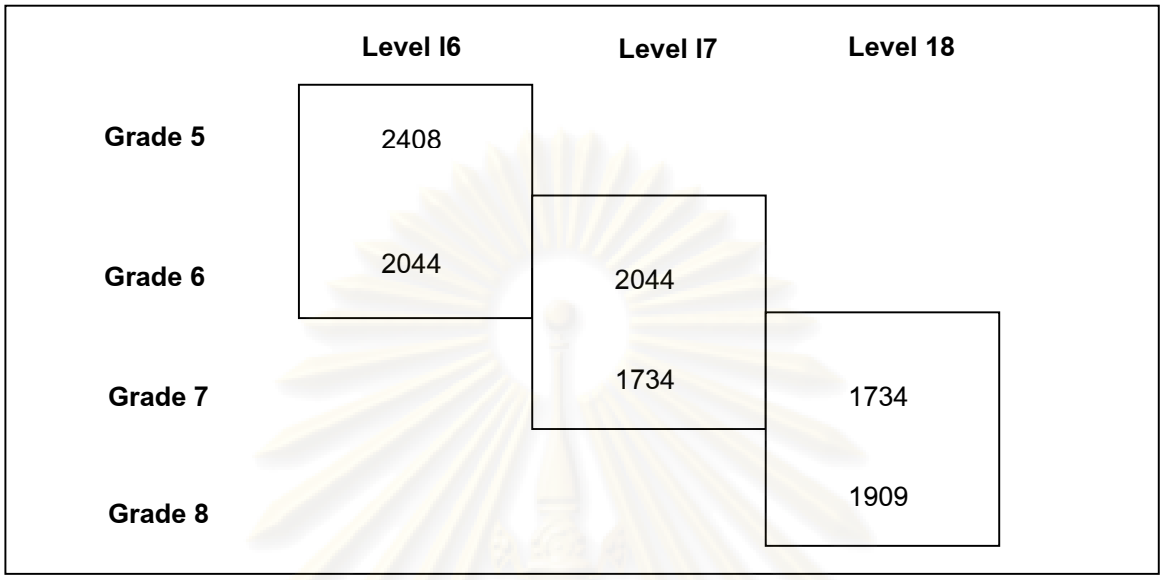
3.2 กระบวนการในการสร้างสเกลแนวตั้งแบบพหุมิติ (Multidimensional Vertical Scaling)

มีประเด็นคำถามจำนวนมากในปัจจุบันว่าจะสามารถนำโมเดล MIRT มาประยุกต์ใช้ศึกษาพัฒนาการแบบปรับเทียบสเกลแนวตั้ง (Vertical Scaling) ได้อย่างไร ซึ่งมีนักวิจัยจำนวนหนึ่งพยามที่จะพัฒนาประเด็นดังกล่าวให้เป็นผลในทางปฏิบัติ เช่น Min (2003) & Reckase & Martineau (2004, 2006) เป็นต้น รวมถึงบทความที่ตีพิมพ์ในวารสาร Educational Measurement: Issues and Practice ในปี 2006 ที่เขียนโดย Ferrara ที่นำมาประยุกต์ใช้ศึกษาพัฒนาการของผลสัมฤทธิ์ทางการเรียนในสาขาวิชาคณิตศาสตร์และวิทยาศาสตร์ของระดับชั้นที่แตกต่างกัน

จากการสังเคราะห์เอกสารและงานวิจัยที่เกี่ยวข้องสามารถสรุปขั้นตอนของกระบวนการสร้างสเกลแนวตั้งแบบพหุมิติ (Multidimensional Vertical Scaling) เพื่อนำไปประยุกต์ใช้ในการศึกษาพัฒนาการ ซึ่งในที่นี้ได้ยกตัวอย่างการออกแบบการเก็บรวบรวมข้อมูลโดยใช้รูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้แบบสอบร่วม ดังแสดงในแผนภาพที่ 2.16 โดยมีขั้นตอนดังนี้

3.2.1 เลือกวิธีการเก็บรวบรวมข้อมูล และเก็บรวบรวมข้อมูลตามรูปแบบที่กำหนดไว้ ซึ่งในที่นี้ได้ยกตัวอย่างการออกแบบการเก็บรวบรวมข้อมูลโดยใช้รูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้แบบสอบร่วม (Anchor-Test Nonequivalent Group Design) ดังแสดงในแผนภาพที่ 2.16

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



แผนภาพที่ 2.16 การออกแบบการเก็บรวบรวมข้อมูลโดยใช้รูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้แบบสอบร่วม (Anchor-Test Nonequivalent Group Design)

จากแผนภาพที่ 2.16 จะเห็นว่านักเรียนในแต่ละระดับจะได้รับแบบสอบ 2 ฉบับที่ต่างกัน เช่น นักเรียนระดับ 6 จะได้รับแบบสอบที่มีความยากในระดับที่ 16 และ 17 ส่วนนักเรียนในระดับที่ 7 จะได้รับแบบสอบที่มีความยากในระดับที่ 17 และ 18 ซึ่งสามารถทำการประมาณค่าพารามิเตอร์ (Calibration) ชุดของข้อสอบโดยใช้โมเดล MIRT ซึ่งชุดของข้อสอบในระดับที่ 17 จะพิจารณาให้เป็นข้อสอบร่วมกับชุดของข้อสอบในระดับที่ 16 และ 18 สรุปว่าแบบสอบในแต่ละระดับจะมี 25 ข้อ ดังนั้นโมเดลที่ใช้สำหรับการประมาณค่าพารามิเตอร์ (Calibration) จะมี 50 ข้อ โดยมีข้อสอบร่วม 25 ข้อ

3.2.2 การประมาณค่าพารามิเตอร์ (Calibration) แบบสอบแต่ละชุดโดยใช้โปรแกรม TESTFACT ที่พัฒนาโดย Bock et al. (2003) หรือ โปรแกรม NOHARM ที่พัฒนาโดย Fraser & McDonald (1986) ซึ่งค่าพารามิเตอร์ของแบบสอบในแต่ละคู่จะถูกหมุนแกน (Rotated) และ ถูกแปลงคะแนนให้อยู่ในสเกลเดียวกัน (Common Configuration) โดยเลือกใช้กระบวนการหมุนแกนแบบตั้งฉาก (Orthogonal Procrustes Rotation) หรือการหมุนแกนแบบไม่ตั้งฉาก (Non-orthogonal Procrustes Rotation) แล้วแต่สถานการณ์ที่ต้องการศึกษา

3.2.3 ภายหลังจากการแปลงคะแนนค่าพารามิเตอร์ของข้อสอบให้มีโครงสร้างร่วมกัน (Coordinate Framework) ในสเปซแบบพหุมิติ (Multidimensional Space) ซึ่งเวคเตอร์ข้อสอบสามารถนำมาพล็อตกราฟเพื่อเปรียบเทียบทักษะและความรู้เพื่อนำมาใช้ในการศึกษาพัฒนาการในแต่ละระดับความสามารถโดยใช้แบบสอบที่ต่างกัน

3.2.4 แปลความหมายของคะแนนที่ได้เพื่อศึกษาการเปลี่ยนแปลงในแต่ละมิติแฝง และศึกษาจากคะแนนรวม (Composite Scores) โดยพิจารณาความแตกต่างของคะแนนในตำแหน่งที่ต่างกันของคะแนนที่ปรับให้อยู่ในสเกลเดียวกัน

3.2.5 ประเมินผลการปรับเทียบสเกล โดยเลือกใช้เกณฑ์การประเมินตามความเหมาะสม (Hariss & Crouse, 1993) เช่น ความเสมอภาค (Weak Equity) ของ Divgi & Yen ที่พิจารณาจากความเท่าเทียมกันของการแจกแจงตามเงื่อนไขของคะแนนที่ได้จากแบบสอบต่างฉบับหลังจากการปรับเทียบแล้ว ดัชนีสำหรับการเปลี่ยนแปลงคะแนน (Indices) ของ Angoff ความคลาดเคลื่อนมาตรฐาน (Standard Error) ของ Angoff เป็นการวิเคราะห์เพื่อประมาณความสามารถของการปรับเทียบจากการสุ่มตัวอย่าง ข้อมูลที่จำลองขึ้น (Generated Data) ของ Lord เป็นการปรับเทียบคะแนนกลับสู่แบบสอบเดิมโดยตรง หรือปรับผ่านแบบสอบอื่นก่อนปรับกลับสู่แบบสอบเดิม กลุ่มตัวอย่างขนาดใหญ่ (Large Sample) ของ Angoff เป็นการใช้กลุ่มตัวอย่างขนาดใหญ่ในการปรับเทียบคะแนนซึ่งคล้ายกับการปรับเทียบคะแนนจากประชากร และใช้เปรียบเทียบกับผลการปรับเทียบกับคะแนนที่มีขนาดกลุ่มตัวอย่างน้อยกว่า ความคงเส้นคงวา (Consistency) เป็นการประเมินผลการปรับเทียบคะแนนข้ามวิธี เพื่อหาความคงเส้นคงวา (Stability) ของ Angoff เป็นการปรับเทียบคะแนนซ้ำเพื่อตรวจความคงที่ของผลการปรับเทียบคะแนน เป็นต้น

3.3 งานวิจัยที่เกี่ยวข้อง

เนื่องจากงานวิจัยที่เกี่ยวข้องกับการการนำโมเดล MIRT ไปประยุกต์ใช้ศึกษาพัฒนาการความสามารถของผู้เรียน ได้กล่าวไว้อย่างชัดเจนในตอนที่ 2 ดังนั้นการนำเสนองานวิจัยที่เกี่ยวข้องในส่วนนี้ผู้วิจัยได้นำเสนอเพื่อให้เห็นความเชื่อมโยงในการพัฒนาวิธีการสร้างสเกลแนวตั้งแบบพหุมิติ (Multidimensional Vertical Scaling) ทั้งต่างประเทศและในประเทศ ดังต่อไปนี้

3.3.1 งานวิจัยที่เกี่ยวข้องในต่างประเทศ

เมื่อพิจารณางานวิจัยที่เกี่ยวข้องในต่างประเทศ ถือได้ว่า Li & Lissitz (2000) เป็นผู้ที่เริ่มทำการศึกษาการปรับเทียบคะแนนในลักษณะของการสร้างสเกลแนวตั้งแบบพหุมิติ (Multidimensional Vertical Equating) โดยอาศัยแนวคิดของโมเดล MIRT โดยเริ่มเห็นผลเป็นรูปธรรมมากกว่าในอดีต แต่ยังเป็นเชิงทดลองจากการจำลองข้อมูล มีข้อจำกัดเพียง 2 มิติเท่านั้น และไม่สามารถระบุได้อย่างชัดเจนว่าเงื่อนไขในการปรับเทียบแนวตั้งใดดีที่สุดได้ ต่อมา Min (2003) ได้ทำการศึกษาการสร้างสเกลแนวตั้งแบบพหุมิติ (Multidimensional Vertical Scaling) ให้เห็นผลเป็นรูปธรรมมากขึ้นโดยเริ่มมีการศึกษาจากข้อมูลจริงและนำไปประยุกต์ใช้กับการศึกษาพัฒนาการ มิติที่ศึกษามีจำนวนมากขึ้น ส่วนการศึกษาของ Min (2003) จะเพิ่มขึ้นเป็น 3 มิติ ซึ่ง Min พบปัญหาสำหรับวิธีการของ Li & Lissitz (2000) ว่าวิธีการปรับเทียบคะแนนตามโมเดล MIRT สำหรับค่าการเลื่อนแกนที่เป็นสเกลลา ไม่เพียงพอสำหรับการศึกษาในหลายมิติ ซึ่งค่าการเลื่อนแกนอาจมีการจะลดหรือขยายสเกลในแต่ละมิติโดยการให้จำนวนที่แน่นอน

เหมือนกัน แต่การแยกการประมาณค่าพารามิเตอร์ (Calibrations) ของแบบสอบหลายชุด อาจจะลดหรือขยายสเกลของการศึกษาในหลายมิติในระดับที่แตกต่างกัน ซึ่งจุดบ่งพร่องนี้ Min (2003) ทำการแทนค่าสเกลลาด้วยเมตริกซ์ (Diagonal Dilation Matrix) เพื่อให้เกิดการเปลี่ยนแปลง สำหรับการศึกษามิติที่มีความหลากหลาย ไม่ใช่การกำหนดค่าเพียงค่าใดค่าหนึ่ง อย่างไรก็ตามจากการศึกษาของ Min (2003) ได้ชี้ให้เห็นจุดอ่อนที่สำคัญคือถ้าจำนวนมิติที่ศึกษามีจำนวนมากขึ้น การคำนวณจะจัดกระทำได้ยากหรือไม่สามารถคำนวณได้ เนื่องจากปฏิสัมพันธ์ระหว่างคุณลักษณะเด่นของมิติข้ามชุดของแบบ

จากที่กล่าวมาจะเห็นว่าวิธีของ Li & Lissitz (2000) & Min (2003) มีความเหมาะสมกับการศึกษาที่จำนวนมิติไม่มาก เช่น ศึกษา 2-3 มิติ เป็นต้น อย่างไรก็ตามเมื่อมีการศึกษาในจำนวนมิติที่มากขึ้น จะมีการพิจารณาความคลาดเคลื่อน ซึ่งผลการศึกษาค่าพารามิเตอร์ความยากของข้อสอบบางครั้งมีความไม่เหมาะสม นักวิจัยจึงพยายามที่จะศึกษาในจำนวนมิติแฝงที่เพิ่มมากขึ้น โดย Reckase & Martineau (2004) ได้ทำการศึกษาวิธีการสร้างสเกลแนวตั้งแบบพหุมิติตามโมเดล MIRT เพื่อนำมาประยุกต์ใช้กับการศึกษาพัฒนาการภาคตัดขวางของผลสัมฤทธิ์ทางการเรียนวิทยาศาสตร์ของนักเรียน 4 ระดับชั้น มีการออกแบบการเก็บรวบรวมข้อมูลรูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้แบบสอบร่วม โดยใช้ข้อสอบร่วมภายใน ระดับละ 25 ข้อ ซึ่งเนื้อหาทางวิทยาศาสตร์ที่ทำการศึกษาพัฒนาการจะมีความซับซ้อนประกอบด้วยการศึกษาทั้งหมด 15 มิติ ทำการประมาณค่าพารามิเตอร์ (Calibration) ข้อมูลในแต่ละระดับชั้นโดยใช้โปรแกรม NOHARM ที่พัฒนาโดย Fraser (1988) ทำการเชื่อมโยงคะแนนในแต่ละระดับชั้นโดยใช้กระบวนการที่พัฒนาโดย Li & Lissitz (2000) & Min (2003) มาดัดแปลงสำหรับการศึกษาในครั้งนี้ โดยใช้เทคนิคการปรับเทียบแบบ MIRT นอกจากนี้มีการวิเคราะห์จัดกลุ่มแบบลดหลั่น (Hierarchical Cluster Analysis) โดยใช้ Ward Method ในการระบุกลุ่มของข้อสอบที่มีลักษณะคล้ายคลึงกันให้อยู่ในกลุ่มเดียวกัน เพื่อพิจารณากลุ่มของข้อสอบที่สามารถจำแนกผลสัมฤทธิ์ทางการเรียนของนักเรียนได้ดีที่สุด ซึ่งผลที่ได้สะท้อนให้เห็นว่า โมเดลแบบพหุมิติเป็นสิ่งจำเป็นที่สะท้อนให้เห็นถึงความซับซ้อนของการศึกษาพัฒนาการภาคตัดขวางของผลสัมฤทธิ์ทางการเรียนวิทยาศาสตร์ได้เป็นอย่างดี แต่ผลการศึกษามีข้อจำกัดของจำนวนข้อสอบร่วมที่มีจำนวนน้อยเกินไป (25 ข้อ) ในการพิจารณาความสัมพันธ์ของข้อสอบที่มีจำนวนไม่มากในแต่ละระดับกับความสัมพันธ์ของเนื้อหา ซึ่งจำเป็นที่จะต้องมีการศึกษาจำนวนของข้อสอบร่วมเพื่อให้การศึกษาพัฒนาการมีความถูกต้องมากยิ่งขึ้น

ต่อมาในปี ค.ศ. 2006 Reckase & Martineau ได้ทำการศึกษาพัฒนาการ โดยเน้นไปที่ความถูกต้องของมิติที่ศึกษา เนื่องจากถึงแม้จะใช้แนวคิดของ MIRT เข้ามาช่วยในการวัดที่ประกอบไปด้วยหลายมิติ แต่การกำหนดมิติที่เหมาะสมยังเป็นเรื่องยากและยังขาดความชัดเจน ซึ่งวิธีการที่พัฒนาขึ้นในการตรวจสอบมิติ เป็นกระบวนการที่เป็นพื้นฐานสำคัญในการสามารถนำไปปฏิบัติได้ง่าย และมีความตรง เป็นขั้นตอนในการสร้างสเกลแนวตั้งเพื่อประเมินข้อมูลที่ได้ โดยเฉพาะงานวิจัยนี้ได้อธิบายการออกแบบวิธีการปรับเทียบคะแนน โดยวิธีที่เรียกว่า “Paired-Grade Scaling Design” เพื่อที่จะทำให้สามารถมั่นใจได้ว่า

มิติที่ศึกษามีความถูกต้องและไม่ผิดรูปไปจากความเป็นจริง ระหว่างการปรับเทียบสเกลจากระดับหนึ่งไปอีกระดับหนึ่ง

ในปีเดียวกัน Yon (2006) ได้ทำการเปรียบเทียบวิธีการเชื่อมโยงคะแนนแบบพหุมิติ 2 วิธี คือ วิธี TCF (Test Characteristic Function) และวิธี NOP (Non-Orthogonal Procrustes) ทั้งจากการจำลองข้อมูลและการใช้ข้อมูลจริงเพื่อนำมาประยุกต์ใช้ในการสร้างสเกลแนวตั้งแบบพหุมิติ (Multidimensional Vertical Scaling) ในการศึกษาพัฒนาการของระดับชั้นที่ต่ำกว่ากับระดับชั้นที่สูงกว่าทั้งหมด 3 ระดับชั้น คือตั้งแต่ระดับที่ 6-8 กับเงื่อนไขที่ใช้ในการศึกษาคือ โครงสร้างมิติความสามารถ 2 ลักษณะ คือ APSS และ MS ซึ่งในปีต่อมา Yon (2007) ได้นำเสนอผลการวิจัยที่มุ่งศึกษาถึงปัญหาการคำนวณความถูกต้องของเมตริกซ์ในการหมุนแกนเมื่อเกิดการเปลี่ยนแปลงโครงสร้าง (Construct Shift) โดยกำหนดให้แบบสอบมีจุดเน้นที่แตกต่างกัน โดยเน้นที่การพิจารณาความแตกต่างระหว่างแบบสอบฐานกับแบบสอบปรับเทียบคะแนน โดยศึกษากับวิธีการเชื่อมโยงคะแนนที่มีลักษณะการหมุนแกนแบบไม่ตั้งฉากตามวิธี NOP เพื่อนำมาประยุกต์ใช้ในการสร้างสเกลแนวตั้งแบบพหุมิติ (Multidimensional Vertical Scaling) ในการศึกษาพัฒนาการในระดับชั้นที่ต่ำกว่ากับระดับชั้นที่สูงกว่า เช่นเดียวกัน

กล่าวโดยสรุป จากการศึกษางานวิจัยของต่างประเทศดังกล่าว เริ่มจากการศึกษาของ Min (2003) Reckase & Martineau (2004, 2006) ต่อเนื่องมาจนถึงการศึกษาของ Yon (2006, 2007) เห็นได้ชัดเจนว่านิยมนำวิธีการเชื่อมโยงคะแนนในบริบทของการปรับเทียบคะแนน ไปประยุกต์ใช้ในการสร้างสเกลแนวตั้งแบบพหุมิติ (Multidimensional Vertical Scaling) เพื่อศึกษาพัฒนาการภาคตัดขวาง ซึ่งกลุ่มผู้สอบมีความสามารถหรือระดับชั้นที่ต่างกัน แต่อยู่ในภาคการศึกษาหรือปีการศึกษาเดียวกัน นั่นเอง

3.3.2 งานวิจัยที่เกี่ยวข้องในประเทศ

ในช่วง 2-3 ปีก่อนหน้านี้จนถึงปัจจุบันแนวโน้มของการศึกษาการเชื่อมโยงคะแนนในรูปแบบของปรับเทียบคะแนนในประเทศไทยจะเน้นไปที่การนำไปประยุกต์ใช้ในการศึกษาพัฒนาการภาคตัดขวาง เนื่องจากเหมาะสมกับบริบทในการพัฒนาลักษณะที่พึงประสงค์ของผู้เรียนในประเทศ โดยเรียกว่า การปรับเทียบแนวตั้ง (Vertical Equating) ซึ่งเหมาะสมสำหรับการปรับเทียบคะแนนในกรณีที่แบบสอบที่มีลักษณะเป็นเอกมิติ แต่ยังไม่มีการศึกษาในบริบทของพหุมิติ โดย ภัทราพร เกษสังข์ (2546) ได้ศึกษาพัฒนาการความสามารถทางคณิตศาสตร์ของนักเรียน และศึกษาประสิทธิภาพของการปรับเทียบแนวตั้งตามโมเดล UIRT แบบโลจิสติก 2 และ 3 พารามิเตอร์ ใช้แบบทดสอบร่วมภายในและแบบทดสอบเชื่อมโยงภายนอก ที่มีความยากแตกต่างกัน (ยากและปานกลาง) และ ความยาว (15 10 และ 5 ข้อ) ของแบบสอบร่วม จากกลุ่มผู้สอบที่มีความสามารถแตกต่างกันสามระดับชั้น คือ ชั้น ม.1,2 และ 3 โดยการวิเคราะห์ความคลาดเคลื่อนมาตรฐานละความเพียงพอของการปรับเทียบคะแนนแต่ละรูปแบบที่ศึกษา แล้วนำรูปแบบการปรับเทียบคะแนนที่เหมาะสม มาใช้ศึกษาพัฒนาการความสามารถทางคณิตศาสตร์ของนักเรียน ผลการวิจัยพบว่า 1) การวิเคราะห์ความคลาดเคลื่อนมาตรฐานของการปรับเทียบคะแนนตามรูปแบบที่ศึกษา

สำหรับผู้ที่มีความสามารถต่างกัน 3 ระดับ ความคลาดเคลื่อนมาตรฐานของการปรับเทียบคะแนนต่ำที่สุดเป็นการปรับเทียบคะแนนในแนวตั้งตามวิธีทฤษฎีตอบสนองข้อสอบ แบบโลจิสติก 2 พารามิเตอร์ ที่แบบทดสอบร่วมภายใน 15 ข้อ มีความยากของข้อสอบร่วมเฉลี่ยอยู่ในระดับยากมาก 2) การวิเคราะห์ความเพียงพอของการปรับเทียบคะแนน พบว่า มีความเพียงพอของการปรับเทียบคะแนนดีที่สุด เป็นการปรับเทียบคะแนนในแนวตั้งตามวิธีทฤษฎีตอบสนองข้อสอบ แบบโลจิสติก 2 พารามิเตอร์ ที่แบบสอบร่วมภายใน 15 ข้อ ที่ลักษณะความยากของข้อสอบร่วมเฉลี่ยระดับยากมาก

ต่อมา พัชรี จันทร์เพ็ง (2547) ได้พัฒนาวิธีการปรับเทียบแนวตั้งตามวิธีทฤษฎีการตอบสนองข้อสอบแบบโลจิสติก 3 พารามิเตอร์ ร่วมกับวิธีการปรับเทียบเชิงเส้นตรงที่ใช้แบบสอบร่วมภายใน เพื่อนำมาใช้ในการศึกษาพัฒนาการความสามารถทางคณิตศาสตร์ของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 1, 2 และ 3 ในแต่ละระดับชั้นโดยภาพรวมและจำแนกตามกลุ่มความสามารถ คือ กลุ่มต่ำ กลุ่มปานกลาง และกลุ่มสูง

จากการศึกษาในประเทศไทยทั้งของ ภัทราพร เกษสังข์ (2546) และ พัชรี จันทร์เพ็ง (2547) ได้ให้ข้อเสนอแนะสำคัญที่ได้จากการวิจัยศึกษาในประเทศคือ ควรมีการพัฒนาวิธีการปรับเทียบคะแนนในกรณีที่แบบสอบมีลักษณะหลายมิติ เนื่องจากสถานการณ์สอบในปัจจุบันเป็นไปได้น้อยที่ข้อสอบหลายฉบับที่นำมาปรับเทียบกันจะมีลักษณะที่มีเนื้อหาเพียงมิติเดียว และควรมีการศึกษาเปรียบเทียบประสิทธิภาพของวิธีการปรับเทียบแนวตั้ง ในหลายวิธี เพื่อนำไปศึกษากับผู้สอบที่มีความสามารถมากกว่า 3 ระดับความสามารถในการนำวิธีการที่ดีที่สุดมาใช้ในการศึกษาความก้าวหน้าหรือพัฒนาการในด้านต่างๆ

กล่าวโดยสรุป จากการสังเคราะห์งานวิจัยทั้งในประเทศและต่างประเทศที่เกี่ยวกับการนำวิธีการเชื่อมโยงคะแนนตามโมเดล MIRT ไปประยุกต์ใช้ในการศึกษาพัฒนาการความสามารถของผู้เรียนในการสร้างสเกลแนวตั้งแบบพหุมิติ (Multidimensional Vertical Scaling) สามารถสรุปประเด็นที่ได้จากการศึกษา ดังนี้

1. วิธีการปรับเทียบแนวตั้ง (Vertical Equating) ตามโมเดล IRT แบบโลจิสติก ชนิด 2 และ 3 พารามิเตอร์ เหมาะสำหรับนำมาใช้ศึกษาในบริบทของการปรับเทียบแนวตั้งแบบเอกมิติ แต่ยังไม่สามารถระบุได้อย่างชัดเจนว่าเหมาะสำหรับนำมาใช้ศึกษาในบริบทของการปรับเทียบแนวตั้งแบบพหุมิติ (Multidimensional Vertical Equating) หรือการสร้างสเกลแนวตั้งแบบพหุมิติ (Multidimensional Vertical Scaling)

2. วิธีการของ Li & Lissitz (2000) & Min (2003) มีความเหมาะสมกับการศึกษาที่จำนวนมิติมีน้อย เช่น ศึกษา 2-3 มิติ เป็นต้น

3. วิธีการเชื่อมโยงคะแนนตามโมเดล MIRT สำหรับค่าการเลื่อนแกนที่เป็นสเกลลา (Scalar Dilation Parameter) ไม่เพียงพอสำหรับการศึกษาในหลายมิติ ที่อาจจะมีการลดหรือขยายสเกลในแต่ละ

มิติ แต่การแยกการประมาณค่าพารามิเตอร์ (Calibration) ของโมเดล MIRT จากแบบสอบหลายๆ ชุด อาจจะลดหรือขยายสเกลของการศึกษาในหลายมิติ ในระดับที่ต่างกัน แนวทางแก้ปัญหาคือแทนค่าการเลื่อนแกนที่เป็นสเกลลา (Scalar Dilation Parameter) ด้วยเมตริกซ์ (Diagonal Dilation Matrix) เพื่อให้เกิดการเปลี่ยนแปลง สำหรับการศึกษามิติที่มีความหลากหลาย ไม่ใช่การกำหนดค่าเพียงค่าใดค่าหนึ่ง (Min, 2003)

4. วิธีการศึกษาของ Reckase & Martineau (2004, 2006) ที่นำวิธีการของ Li & Lissitz (2000) & Min (2003) มาประยุกต์ใช้ประกอบกับการหมุนแกนการแปลงคะแนนแบบไม่ตั้งฉาก (Non-Orthogonal Procrustes Transformation) (Mulaik, 1972 อ้างใน Reckase & Martineau, 2004) จะทำการหมุนแกนเป็นไปโดยอัตโนมัติในการแปลงคะแนนในแต่ละมิติของเมตริกซ์ปรับเทียบคะแนนกับมิติของเมตริกซ์พื้นฐาน ซึ่งขึ้นอยู่กับความสอดคล้องที่ดีที่สุดของเมตริกซ์ปรับเทียบคะแนนกับเมตริกซ์ฐาน ขณะเดียวกันการแปลงคะแนนแบบไม่ตั้งฉาก (Non-Orthogonal) จะไม่นำค่าพารามิเตอร์การเลื่อนแกน มาคำนวณเป็นการช่วยลดความยุ่งยากได้ในระดับหนึ่ง

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง ที่ผู้วิจัยได้นำเสนอทั้ง 3 ตอน คือ ตอนที่ 1 โมเดล MIRT ตอนที่ 2 การเชื่อมโยงคะแนนตามโมเดล MIRT และ ตอนที่ 3 การสร้างสเกลแนวตั้งแบบพหุมิติ (Multidimensional Vertical Scaling) รวมถึงงานวิจัยที่เกี่ยวข้อง สามารถสังเคราะห์ประเด็นที่ได้จากงานวิจัยที่เกี่ยวข้องและข้อเสนอแนะเพื่อนำไปศึกษาวิจัยในอนาคต ดังนี้

1. ประเด็นที่สังเคราะห์ได้จากงานวิจัยที่เกี่ยวข้อง

ข้อค้นพบที่ได้จากการศึกษาการปรับเทียบคะแนนตามโมเดล MIRT สามารถสรุปได้ตามองค์ประกอบที่เกี่ยวข้องกับการเชื่อมโยงคะแนนแบบพหุมิติ ดังนี้

1.1 ชนิดของโมเดลที่ทำการศึกษา (MIRT Linking Model) จะเห็นได้อย่างชัดเจนว่าโมเดลโลจิสติกแบบพหุมิติ ชนิด 2 พารามิเตอร์ (M2PL) และ 3 พารามิเตอร์ (M3PL) นิยมนำมาใช้ในการเชื่อมโยงคะแนนตามโมเดล MIRT (Lissitz, 2000; Reckase & Martineau, 2004; Reckase & Martineau, 2006) แต่ยังไม่สามารถระบุได้ชัดเจนว่า โมเดลโลจิสติกชนิดใดเหมาะสำหรับนำมาศึกษาในบริบทของการปรับเทียบแนวนอนแบบพหุมิติ (Multidimensional Horizontal Equating) หรือการปรับเทียบแนวตั้งแบบพหุมิติ (Multidimensional Vertical Equating) รวมถึงการสร้างสเกลแนวตั้งแบบพหุมิติ (Multidimensional Vertical Scaling) นอกจากนี้เมื่อพิจารณาถึงความเหมาะสมในการนำโมเดล MIRT มาใช้ศึกษาในบริบทของ Multidimensional Vertical Equating/Scaling พบว่าจากการศึกษาที่ผ่านมานิยมนำโมเดลแบบโลจิสติกชนิด 2 และ 3 พารามิเตอร์ สำหรับนำมาใช้ศึกษาในบริบทของ Unidimensional Vertical Equating แต่ยังไม่สามารถระบุได้อย่างชัดเจนว่าเหมาะสำหรับนำมาใช้ศึกษาในบริบทของ Multidimensional Vertical Equating/Scaling หรือไม่อย่างไร

1.2 โมเดล MIRT (MIRT Model) โดยทั่วไปโมเดลที่นิยมมากที่สุดในการศึกษาการเชื่อมโยงคะแนน คือ โมเดลที่การประมาณค่าความสามารถ (θ) ในมิติต่างกันสามารถชดเชยกันได้ (Compensatory Model) เนื่องจากไม่ยุ่งยากซับซ้อนและเห็นผลในทางปฏิบัติ ประกอบกับ โมเดลที่การประมาณค่าความสามารถ (θ) ในมิติต่างกันไม่สามารถชดเชยกันได้ (Non-Compensatory Model) มีความยุ่งยากในทางปฏิบัติ ซึ่งขั้นตอนในการคำนวณไม่มีประสิทธิภาพที่เพียงพอในการประมาณค่าพารามิเตอร์

1.3 แบบแผนการเชื่อมโยงคะแนน (Linking Design) จากการศึกษาพบว่าแบบแผนแบบสอบร่วม มีความเป็นไปได้และเป็นที่นิยมในทางปฏิบัติ ซึ่งการศึกษาส่วนมากกำหนดและอภิปรายในแบบแผนนี้ (Oshima & Davey,1994; Li & Lissitz, 2000; Briggs, 2003; Min, 2003; Reckase & Martineau, 2004; Reckase & Martineau, 2006) ซึ่งไม่มีความซับซ้อนในการจัดการกับแบบสอบหลายๆ ฉบับ โดยการศึกษาที่ผ่านมาพบว่า Reckase & Martineau (2004, 2006) ได้ออกแบบการเก็บรวบรวมข้อมูลรูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้แบบสอบร่วม (Anchor-Test Nonequivalent Group Design) ข้อสอบร่วมภายใน (Common Items) ระดับละ 25 ข้อ ซึ่งข้อค้นพบจากการศึกษาดังกล่าวพบว่าใช้ข้อสอบร่วมภายในน้อยเกินไปกับกลุ่มผู้สอบที่มีขนาดใหญ่ อย่างไรก็ตามประเด็นในเรื่องการใช้ข้อสอบร่วมจำนวนเท่าใดจึงจะเพียงพอต่อการปรับเทียบคะแนนตามแนว MIRT ยังไม่มีความชัดเจนในประเด็นนี้ซึ่งต้องมีการศึกษาต่อไปในอนาคต

1.4 ความยาวแบบสอบ (Test Length) แต่ละการศึกษาใช้แบบสอบที่มีความยาวแตกต่างกันจึงยังไม่มีข้อสรุปที่ชัดเจนเกี่ยวกับความยาวของแบบสอบ เช่น Oshima, Davey, & Lee (2000) ใช้จำนวนข้อสอบ 40 ข้อ หรือ Reckase & Martineau (2004, 2006) ศึกษากับความยาวแบบสอบ 50 ข้อ เป็นต้น

1.5 ความยาวของข้อสอบร่วม (Anchor Length) เช่นเดียวกัน มีความแตกต่างในแต่ละการศึกษา ซึ่งผลที่ได้อาจต้องตีความอย่างรอบคอบ อย่างไรก็ตามจำนวนข้อสอบร่วมที่ยาวกว่าย่อมให้ผลที่ดีกว่า ซึ่งจากการศึกษาของ Li & Lissitz (2000) ที่ศึกษาจำนวนของข้อสอบร่วม 15 ข้อ และ 25 ข้อ และ Reckase & Martineau (2004, 2006) ได้การออกแบบการเก็บรวบรวมข้อมูลรูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้แบบสอบร่วม (Anchor-Test Nonequivalent Group Design) โดยใช้ข้อสอบร่วมภายใน ระดับละ 25 ข้อ ข้อค้นพบจากการศึกษาดังกล่าวพบว่าใช้ข้อสอบร่วมภายในที่น้อยเกินไปกับกลุ่มผู้สอบที่มีขนาดใหญ่ อย่างไรก็ตามประเด็นในเรื่องการใช้ข้อสอบร่วมจำนวนเท่าใดจึงจะเพียงพอต่อการปรับเทียบคะแนนตามแนว MIRT ยังไม่มีความชัดเจนในประเด็นนี้ซึ่งต้องมีการศึกษาต่อไปในอนาคต

1.6 ขนาดของหน่วยตัวอย่าง (Sample Size) ไม่มีการตกลงเกี่ยวกับประเด็นนี้ แต่ละการศึกษาประยุกต์ใช้หน่วยตัวอย่างขนาดแตกต่างกัน เช่น Li & Lissitz (1997, 2000) ได้ทำการศึกษาขนาดของกลุ่มตัวอย่างขั้นต่ำที่ส่งผลต่อการปรับเทียบคะแนนตามแนว MIRT โดยศึกษาจากกลุ่มตัวอย่างจำนวน 1,000 คน 2,000 คน และ 3,000 คน ภายใต้เงื่อนไขการศึกษาประสิทธิภาพการปรับเทียบคะแนนที่แตกต่างกัน ซึ่งผลการศึกษาพบว่ากลุ่มตัวอย่างที่ใช้ในการปรับเทียบคะแนนแต่ละกลุ่มไม่ควรต่ำกว่า 2,000 คน โดยทั่วไปแล้วหน่วยตัวอย่างขนาดใหญ่จะทำให้เกิดความแม่นยำสำหรับการประมาณค่าพารามิเตอร์ของข้อสอบ

1.7 จำนวนมิติแฝงที่ทำการศึกษา (m-Dimensional) ยังไม่สามารถระบุได้ว่าควรจะมีจำนวนเท่าใดจึงจะเพียงพอต่อนำมาศึกษา โดยมีการศึกษาในจำนวนมิติแฝงที่แตกต่างไปตามบริบท เช่น มีการศึกษา 2 มิติ (Oshima & Davey, 1994; Oshima, Davey, & Lee, 1997; Oshima, Davey, & Lee, 2000; Li & Lissitz, 2000) การศึกษา 3 มิติ (Min, 2003) การศึกษา 4 มิติ (Briggs, 2003) การศึกษา 5 มิติ และ การศึกษา 15 มิติ (Reckase & Martineau, 2004; Reckase & Martineau, 2006) อย่างไรก็ตามจากการศึกษาพบว่า วิธีการของ Li & Lissitz (2000) & Min (2003) มีความเหมาะสมกับการศึกษาที่จำนวนมิติ 2-3 มิติ เท่านั้น

1.8 เทคนิควิธีที่ใช้ในการปรับเทียบคะแนน พบว่านิยมศึกษา 4 วิธี คือ 1) Orthogonal Procrustes Rotation 2) Non-orthogonal Procrustes Rotation 3) Translation Tranformation และ 4) Single Dilation เพื่อนำไปสู่การประมาณค่า 3 ค่า คือ 1) Othogonal Procruster Rotation Matrix (T) 2) Diagonal Dilation Matrix (K) และ 3) Translation Vector (m) ซึ่งเป็นส่วนประกอบที่สำคัญในการแปลงคะแนนองค์ประกอบ (Composite Transformation) เพื่อปรับเทียบคะแนนจากแบบสอบปรับเทียบคะแนน (Equated Form) ไปยังแบบสอบฐาน (Based Form) โดยมีวิธีหรือตัวที่ใช้ในการประมาณค่า (Estimators) ประกอบด้วย 1) ประมาณค่า k โดยวิธี Ratio of Eigenvalues, Trace, Least Square และ MTRs 2) ประมาณค่าพารามิเตอร์ในมิติของ m_n โดยวิธี Least Square และ MTRs (Oshima & Davey,1994; Li & Lissitz, 2000; Briggs, 2003; Min, 2003; Reckase & Martineau, 2004; Reckase & Martineau, 2006)

นอกจากนี้จากการศึกษาของ Reckase & Martineau (2006) พบว่า การศึกษาจากเวคเตอร์ (Vector Approach) เหมาะสำหรับตรวจสอบมิติที่มีจำนวนมิติมากและจำนวนตัวแปรมากทั้งองค์ประกอบที่มีความสัมพันธ์กันและองค์ประกอบที่ไม่มีความสัมพันธ์กัน ซึ่งตรงข้ามกับการศึกษาด้วย Scree Plot ที่เมื่อศึกษากับมิติเพิ่มขึ้นจะทำให้ความแม่นยำลดลง ส่วนผลการศึกษาจากข้อมูลจริง ที่มีข้อจำกัดเพียง 3 มิติ จากการวิเคราะห์ EFA ด้วยการหมุนแกนแบบ Varimax และ Promax มีความเหมาะสมส่วนศึกษาด้วยเวคเตอร์ (Vector Approach)

1.9 การแจกแจงความสามารถ (Ability Distribution) เมื่อนำแบบสอบหลายฉบับดำเนินการต่อกลุ่มสุ่มหรือกลุ่มที่สมมูลกัน ผลที่ได้จะให้เกิดความแม่นยำมากกว่า เมื่อแบบสอบหลายฉบับดำเนินการในกลุ่มที่ไม่เท่าเทียมกัน แต่ถ้าต้องการปรับเทียบคะแนนในกลุ่มที่ไม่เท่าเทียมกันควรใช้แบบแผนแบบสอบร่วมจะเหมาะสมกว่า (Thompson et al., 1997)

1.10 โครงสร้างมิติความสามารถ (Dimensional Structure) โดยทั่วไปโครงสร้างมิติความสามารถจะประกอบไปด้วย 3 ลักษณะ คือ โครงสร้างอย่างง่าย (Simple Structure; SS) โครงสร้างที่ไม่ซับซ้อน (Approximate Simple Structure; APSS) และ โครงสร้างที่ซับซ้อนมากขึ้น (Mixed Structure; MS) แต่สถานการณ์โดยทั่วไปนิยมศึกษาโครงสร้างมิติความสามารถเพียง 2 ลักษณะ คือ โครงสร้างที่ไม่ซับซ้อนและโครงสร้างที่ซับซ้อนมากขึ้น เนื่องจากสอดคล้องกับสถานการณ์จริงมากกว่าโครงสร้างอย่างง่าย (Roussos, Stout, & Marden, 1998; Min, 2003; Yon, 2006)

1.11 การประเมินประสิทธิภาพการปรับเทียบคะแนน โดยส่วนใหญ่นิยมใช้ค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (Root Mean Square Error; RMSE) และค่าความลำเอียง (BIAS) เป็นดัชนีในการประเมินประสิทธิภาพการปรับเทียบคะแนน (Hirsch, 1989; Oshima & Davey, 1994; Oshima, Davey, & Lee, 2000; Li & Lissitz, 2000) เช่น จากการศึกษาของ Oshima & Davey (1994) มีการประเมินความคลาดเคลื่อนของการปรับเทียบคะแนนโดยการเปรียบเทียบความแตกต่างของค่าเฉลี่ย (Mean Difference) และการเปรียบเทียบค่าสัมบูรณ์ความแตกต่างของค่าเฉลี่ย (Mean Absolute Difference) ระหว่างความสามารถที่แท้จริงกับความสามารถที่สังเกตได้ เกณฑ์ที่ใช้ประเมินประสิทธิภาพจะศึกษาจากความคลาดเคลื่อน RMSE เป็นต้น

1.12 การนำไปประยุกต์ใช้ (Application) ในวงการศึกษามีการนำวิธีเชื่อมโยงคะแนนตามโมเดล MIRT ในบริบทของการปรับเทียบคะแนน ไปใช้ในการเชื่อมโยงคะแนนแบบสอบต่างฉบับที่วัดคุณลักษณะเดียวกัน และใช้ในการศึกษาพัฒนาการหรือความก้าวหน้าทางการทางการเรียน ซึ่งจากการศึกษาพัฒนาการทางการเรียน ทั้งในประเทศและต่างประเทศ มีลักษณะของการสร้างสเกลแนวตั้งแบบพหุมิติ (Multidimensional Vertical Scaling) โดยเน้นไปที่การนำไปประยุกต์ใช้ได้จริงในทางปฏิบัติมากกว่าการทดลองเหมือนในอดีต เช่นจากการศึกษาและพัฒนาอย่างต่อเนื่องของ Reckase & Martineau (2004, 2006) ตลอดจนการศึกษาของ Yon (2006, 2007)

ดังได้กล่าวมาแล้วข้างต้น โดยภาพรวมข้อค้นพบที่ได้จากการศึกษาการเชื่อมโยงคะแนนแบบพหุมิติ อาจจะสามารถอ้างอิงหรือไม่สามารถอ้างอิงไปยังสถานการณ์และเงื่อนไขอื่นได้ เพราะว่าแต่ละการศึกษาจะเกี่ยวข้องกับบริบทเฉพาะ ผลสรุปอาจจะเหมือนหรือไม่เหมือนกับการศึกษาอื่น ๆ นอกเหนือจากนี้ บางการศึกษาใช้ข้อมูลจริง และบางการศึกษาอาจจำลองข้อมูล ทำให้ผลของการศึกษาที่ได้มีความแตกต่างกัน ซึ่งจากการศึกษาที่ผ่านมาพบว่า ไม่สามารถระบุได้อย่างชัดเจนว่าเงื่อนไขใดทำ

ให้การเชื่อมโยงคะแนนในแต่ละวิธีให้ผลที่แม่นยำเพียงใด จึงมีความจำเป็นที่จะต้องมีการศึกษาวิจัยเพื่อขยายองค์ความรู้ทางด้านการเชื่อมโยงคะแนนแบบพหุมิติตามโมเดล MIRT โดยมีข้อเสนอแนะเพื่อนำไปศึกษาวิจัยในอนาคต ดังที่จะกล่าวถึงในรายละเอียดต่อไป

2. ข้อเสนอแนะเพื่อนำไปศึกษาวิจัยในอนาคต

2.1 จากการศึกษางานวิจัยที่เกี่ยวข้องในประเทศ (ภัทราพร เกษสังข์, 2546; พัชรี จันทร์เพ็ง, 2547) มีข้อเสนอแนะที่สำคัญ 2 ประเด็นย่อย คือ

(1) ควรมีการพัฒนาวิธีการเชื่อมโยงคะแนนในกรณีที่แบบสอบที่มีลักษณะหลายมิติ เนื่องจากสถานการณ์สอบในปัจจุบันเป็นไปได้น้อยที่ข้อสอบหลายฉบับที่นำมาปรับเทียบกันจะมีลักษณะที่มีเนื้อหาเพียงมิติเดียว ซึ่งวัดคุณลักษณะเด่นเพียงคุณลักษณะเดียว

(2) ควรมีการศึกษาเปรียบเทียบประสิทธิภาพของวิธีการปรับเทียบแนวตั้งแบบพหุมิติ (Multidimensional Vertical Equating) ในหลายๆ วิธี เพื่อนำไปศึกษากับผู้สอบที่มีความสามารถมากกว่า 3 ระดับความสามารถเพื่อนำวิธีการที่ดีที่สุดมาใช้ในการศึกษาความก้าวหน้าหรือพัฒนาการในด้านต่างๆ

2.2 จากการศึกษางานวิจัยที่เกี่ยวข้องในต่างประเทศ (Hirsch, 1989; Oshima, & Davey,1994; Davey & Lee, 1997; Oshima, Davey, & Lee, 2000; Li & Lissitz, 2000; Briggs, 2003; Min, 2003; Reckase & Martineau, 2004; Reckase & Martineau, 2006) มีข้อเสนอแนะที่สำคัญ คือ

(1) ควรมีการศึกษาเปรียบเทียบประสิทธิภาพของวิธีการเชื่อมโยงคะแนนตามโมเดล MIRT ในหลายๆ เงื่อนไข เช่น

(1.1) โมเดลการเชื่อมโยงคะแนน
(1.2) วิธีการเชื่อมโยงคะแนน
(1.3) โครงสร้างมิติความสามารถ
(1.4) รูปแบบการเชื่อมโยงคะแนน
(1.5) เทคนิคการเชื่อมโยงคะแนน
(1.6) ตัวประมาณค่าพารามิเตอร์
(1.7) จำนวนมิติแฝง
(1.8) จำนวนข้อสอบร่วม
(1.9) ขนาดของกลุ่มตัวอย่าง

จากการศึกษาที่ผ่านมา พบว่าควรให้ความสำคัญกับจำนวนมิติที่ศึกษา เนื่องจากการกำหนดมิติที่เหมาะสมยังเป็นเรื่องยากและยังขาดความชัดเจนประกอบกับผลการศึกษามีข้อจำกัดของจำนวนข้อสอบร่วมที่มีจำนวนน้อยเกินไป (25 ข้อ) ในการที่จะดูความสัมพันธ์ของข้อสอบที่มีจำนวนไม่

มากในแต่ละระดับกับความสัมพันธ์ของเนื้อหา ซึ่งจำเป็นที่จะต้องมีการศึกษาจำนวนของข้อสอบร่วมเพื่อให้การศึกษาพัฒนาการมีความถูกต้องมากยิ่งขึ้น

(2) ควรพัฒนาวิธีการตรวจสอบมิติแฝงที่ทำการศึกษาให้มีความชัดเจนมากยิ่งขึ้น เนื่องจากการตรวจสอบมิติ เป็นกระบวนการขั้นพื้นฐานสำคัญที่สามารถนำไปปฏิบัติได้ง่าย และมีความตรง และเป็นขั้นตอนในการเชื่อมโยงคะแนนแนวตั้งเพื่อประเมินข้อมูลที่ได้ โดยถ้าสามารถประเมินมิติได้อย่างถูกต้อง ข้อมูลเชิงประจักษ์ที่ได้สามารถปรับเทียบข้ามระดับความสามารถโดยไม่เกี่ยวข้องกับกระบวนการทำงานที่ใช้เวลาในการตั้งชื่อมิติ

ดังนั้นจากการศึกษาเอกสารและงานวิจัยทั้งหมด ผู้วิจัยจึงสนใจเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนน 2 วิธี คือ วิธีการเชื่อมโยงคะแนนตามวิธีของ Min (2003) เรียกว่า M (Min's Method) และวิธีการเชื่อมโยงคะแนนของ Reckase & Martineau (2004) เรียกว่า NOP (Non-Orthogonal Procrustes) ซึ่งทั้งสองวิธีมีลักษณะการหมุนแกนที่ต่างกัน ภายใต้เงื่อนไขที่ทำการศึกษาคือ โครงสร้างมิติความสามารถที่แตกต่างกัน 2 ลักษณะ คือโครงสร้างที่ไม่ซับซ้อน (APSS) และโครงสร้างที่ซับซ้อนมากขึ้น (MS) และระดับความสัมพันธ์ของมิติความสามารถที่แตกต่างกัน 3 ระดับ คือ ไม่มีความสัมพันธ์กัน (r=0) มีความสัมพันธ์กันในระดับกลาง (r=0.5) และมีความสัมพันธ์กันในระดับสูง (r=0.8) เพื่อประโยชน์ในการนำไปเลือกใช้กับสถานการณ์ต่าง ๆ ให้เหมาะสมกับข้อมูลจริงต่อไปในอนาคต ในการแก้ปัญหาการจัดการศึกษาในประเทศไทยปัจจุบันที่จำเป็นจะต้องมีการวิจัยเชิงพัฒนาการของนักเรียนในระดับชั้นต่าง ๆ อย่างต่อเนื่อง ในการนำไปใช้ปรับปรุงแก้ไขได้ทันท่วงที เพื่อให้สอดคล้องกับการพัฒนาผู้เรียนที่สอดคล้องกับแผนการศึกษาของชาติและพระราชบัญญัติการศึกษาแห่งชาติ พ.ศ. 2542 และแก้ไขเพิ่มเติม (ฉบับที่ 2) พ.ศ. 2545 ที่มุ่งเน้นการประเมินความก้าวหน้าการเรียนรู้ของผู้เรียน ดังนั้นผู้วิจัยจึงสนใจศึกษาวิธีการเชื่อมโยงแบบพหุมิติตามโมเดล MIRT ในรูปแบบของการปรับเทียบคะแนนไปประยุกต์ใช้ในบริบทของการสร้างสเกลแนวตั้ง (Multidimensional Vertical Scaling) ซึ่งเป็นวิธีการเชื่อมโยงคะแนนที่ได้รับความนิยมมากที่สุดในต่างประเทศ เพื่อนำมาใช้ในการศึกษาพัฒนาการความสามารถของผู้เรียนที่มีอยู่ในระดับชั้นที่แตกต่างกันทั้งในบริบทของการวัดผลสัมฤทธิ์ของผู้เรียน รวมถึงการวัดความถนัดด้วยเช่นกัน ซึ่งเป็นรูปแบบวิธีการเชื่อมโยงคะแนนที่มีความยืดหยุ่นและสอดคล้องกับสถานการณ์การทดสอบมากกว่า การปรับเทียบ

แนวตั้งแบบพหุมิติ (Multidimensional Vertical Equating) (Kolen & Brennan, 2004) โดยศึกษาพัฒนาการในช่วงชั้นใดช่วงชั้นหนึ่ง คือ ศึกษาใน 3 ระดับชั้นที่แตกต่างกัน

สำหรับเกณฑ์ที่ใช้ในการเปรียบคุณภาพในครั้งนี้ ผู้วิจัยเลือกใช้ดัชนีค่าความลำเอียง (BIAS) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (Root-Mean Square Error; RMSE) ซึ่งเป็นดัชนีบ่งชี้คุณภาพของวิธีการเชื่อมโยงคะแนนใน 2 วิธี คือ วิธี M และ วิธี NOP ภายใต้เงื่อนไขที่ทำการศึกษา คือ โครงสร้างมิติความสามารถและระดับความสัมพันธ์ของมิติความสามารถที่แตกต่างกัน เนื่องจากการใช้ดัชนีเหล่านี้เป็นเกณฑ์ในการเชื่อมโยงคะแนนสามารถแปลความหมายตามขนาดดัชนี ซึ่งเป็นค่าที่เรียงลำดับได้ว่า ดัชนีที่มีค่าน้อยแสดงว่าความแตกต่างนั้นมีน้อย ซึ่งค่าดัชนีที่ได้มีลักษณะเป็นค่ามาตรฐาน เพราะถ่วงน้ำหนักด้วยความแปรปรวนของคะแนนเกณฑ์ สามารถนำมาเปรียบเทียบกันได้โดยตรง ถึงแม้ในสถานการณ์ที่ได้ข้อมูลต่างกันก็ตาม ซึ่งผลที่ได้จากการศึกษาจะเป็นเป็นสารสนเทศที่สำคัญในการเลือกใช้วิธีการเชื่อมโยงคะแนนที่เหมาะสมและมีความถูกต้อง นอกจากนี้มีการพิจารณาค่าความสัมพันธ์ระหว่างคะแนนที่แปลงได้จากแบบสอบฐานกับคะแนนเกณฑ์ที่ประมาณได้จากแบบสอบฐาน (CORR) พร้อมทั้งวิเคราะห์ความแปรปรวนแบบพหุจำแนก 3 ทาง (Three-way MANOVA) ของค่าเฉลี่ยดัชนี BIAS, RMSE และ CORR สำหรับศึกษาปฏิสัมพันธ์ของวิธีการเชื่อมโยงคะแนนและเงื่อนไขที่ทำการศึกษาคือโครงสร้างมิติความสามารถและระดับความสัมพันธ์ของมิติความสามารถ ซึ่งผลที่ได้จากการวิจัยในครั้งนี้จะเป็นประโยชน์ในการนำไปใช้ประกอบการตัดสินใจ ในการเลือกใช้วิธีการเชื่อมโยงคะแนนที่เหมาะสมและสอดคล้องกับสถานการณ์ที่ต้องการศึกษามากที่สุด

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 3
วิธีดำเนินการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนน 2 วิธี ระหว่างวิธีการหมุนแกนแบบตั้งฉากโดยวิธี M (Min, 2003) และวิธีการหมุนแกนแบบไม่ตั้งฉากโดยวิธี NOP (Non-Orthogonal Procrustes) (Reckase & Martineau, 2004) ภายใต้เงื่อนไขที่ทำการศึกษาคือ โครงสร้างมิติความสามารถ (Dimensional Structure) ที่ต่างกัน 2 ลักษณะ และระดับความสัมพันธ์ของมิติความสามารถ 3 ระดับ ซึ่งศึกษาจากการจำลอง (Simulation Study) เกณฑ์ที่ใช้ในการเปรียบเทียบคุณภาพวิธีการเชื่อมโยงคะแนนทั้ง 2 วิธี คือ พิจารณาจากดัชนีค่าความลำเอียง (BIAS) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (Root-Mean Square Error; RMSE) พร้อมทั้งศึกษาค่าความสัมพันธ์ระหว่างคะแนนที่แปลงได้กับคะแนนเกณฑ์ (CORR) ของผลที่ได้จากการแปลงคะแนนระหว่างแบบสอบฐาน และแบบสอบปรับเทียบคะแนน นอกจากนี้ผู้วิจัยทำการวิเคราะห์ความแปรปรวนแบบพหุจำแนก 3 ทาง (Three-way MANOVA) เพื่อเปรียบเทียบค่าเฉลี่ยของ BIAS, RMSE และ CORR สำหรับศึกษาปฏิสัมพันธ์ของวิธีการเชื่อมโยงคะแนนและเงื่อนไขที่ทำการศึกษา ซึ่งผลที่ได้จากการวิจัยครั้งนี้จะเป็นประโยชน์เพื่อประกอบการตัดสินใจในการเลือกใช้วิธีการเชื่อมโยงคะแนนที่เหมาะสมและสอดคล้องกับสถานการณ์ที่ต้องการศึกษามากที่สุด เพื่อประโยชน์ในการนำไปเลือกใช้กับสถานการณ์ต่าง ๆ ให้เหมาะสมกับข้อมูลจริงต่อไปในอนาคต โดยมีขั้นตอนการวิจัยตามลำดับ ดังนี้

1. ศึกษาค้นคว้าเกี่ยวกับทฤษฎีและหลักการพร้อมทั้งวิธีการที่ใช้ในการเชื่อมโยงคะแนนแบบพหุมิติตามโมเดล MIRT และงานวิจัยที่เกี่ยวข้อง จากเอกสาร หนังสือ วารสารและรายงานการวิจัยด้านการวัดผลการศึกษา ทั้งในประเทศและต่างประเทศ

2. ศึกษารายละเอียดการใช้โปรแกรมสำเร็จรูป MATLAB เพื่อใช้ในการจำลองข้อมูลตามเงื่อนไขต่าง ๆ ที่ต้องการศึกษา และโปรแกรม TESTFACT เพื่อใช้ในการประมาณค่าพารามิเตอร์แบบพหุมิติ

3. จำลองข้อมูลการตอบข้อสอบของผู้ตอบตามแบบแผนและตัวแปรที่กำหนด ดังแสดงในแผนภาพที่ 3.1

4. วิเคราะห์ข้อมูล และเปรียบเทียบคุณภาพของการเชื่อมโยงคะแนนทั้ง 2 วิธี คือวิธี M และวิธี NOP เมื่อมีโครงสร้างมิติความสามารถแตกต่างกัน 2 ลักษณะ คือโครงสร้างที่ไม่ซับซ้อน (APSS) และโครงสร้างที่ซับซ้อนมากขึ้น (MS) และมีระดับความสัมพันธ์ของมิติความสามารถแตกต่างกัน 3 ระดับ โดยพิจารณาจากค่า BIAS และ RMSE ซึ่งเป็นดัชนีในการประเมินคุณภาพการเชื่อมโยงคะแนน พร้อมทั้งศึกษาค่าสัมประสิทธิ์สหสัมพันธ์ภายหลังจากการแปลงคะแนนของแบบสอบปรับเทียบคะแนนกับแบบสอบฐาน พร้อมทั้งวิเคราะห์ความแปรปรวนแบบพหุ เพื่อเปรียบเทียบค่าเฉลี่ยของ RMSE BIAS และ CORR สำหรับศึกษาปฏิสัมพันธ์ของวิธีการเชื่อมโยงคะแนน โครงสร้างมิติความสามารถ และระดับความสัมพันธ์ระหว่างมิติความสามารถ

5. สรุปผลการวิเคราะห์ข้อมูล เพื่อเปรียบเทียบคุณภาพของการเชื่อมโยงคะแนนในแต่ละเงื่อนไขที่ทำการศึกษา

ตัวแปรที่ใช้ในการวิจัย

เนื่องจากการวิจัยครั้งนี้มุ่งเปรียบเทียบประสิทธิภาพของวิธีการเชื่อมโยงคะแนน 2 วิธี ในเงื่อนไขที่ทำการศึกษา คือโครงสร้างมิติความสามารถที่แตกต่างกัน 2 ลักษณะ และระดับความสัมพันธ์ของมิติความสามารถที่แตกต่างกัน 3 ระดับ โดยมุ่งศึกษากับโมเดลแบบพหุมิติ 3 พารามิเตอร์ (M3PL) โดยมีรายละเอียดดังต่อไปนี้

ตัวแปรอิสระ คือ **วิธีการเชื่อมโยงคะแนน (Linking Method)** มี 2 วิธี คือ

1.1 การหมุนแกนแบบตั้งฉากโดยวิธี M (Min's Method)

1.2 การหมุนแกนแบบไม่ตั้งฉากโดยวิธี NOP (Non-Orthogonal Procrustes Method)

ประกอบด้วยเงื่อนไขที่ทำการศึกษา คือ

1) โครงสร้างมิติความสามารถ (Dimensional Structure) มี 2 ลักษณะ คือ

1.1) โครงสร้างที่ไม่ซับซ้อน (Approximate Simple Structure; APSS)

1.2) โครงสร้างที่ซับซ้อน (Mixed Structure; MS)

2) ระดับความสัมพันธ์ของมิติความสามารถ (Correlation Between θ_1 & θ_2) มี 3 ระดับ คือ

2.1) ไม่มีความสัมพันธ์ (r=0)

2.2) ความสัมพันธ์ระดับกลาง (r=0.5)

2.3) ความสัมพันธ์ระดับสูง (r=0.8)

ตัวแปรตาม คือ คุณภาพของการเชื่อมโยงคะแนน (Quality of Linking Method) ตามโมเดล MIRT

จากที่กล่าวมาข้างต้นเป็นการสรุปขั้นตอนดำเนินงานวิจัยโดยรวมตั้งแต่การศึกษาค้นคว้าเอกสารและงานวิจัยที่เกี่ยวข้องจนถึงการวิเคราะห์และสรุปผลการวิจัย รวมถึงตัวแปรที่ใช้ในการวิจัย ในรายละเอียดต่อไป จะนำเสนอวิธีดำเนินการวิจัยโดยละเอียด ซึ่งในที่นี้แบ่งออกเป็น 3 ตอน ได้แก่

ตอนที่ 1 เงื่อนไขที่ใช้ในการจำลองข้อมูล

ตอนที่ 2 การศึกษาการจำลองข้อมูล

ตอนที่ 3 การวิเคราะห์ข้อมูลที่ได้จากการจำลองข้อมูลและเกณฑ์ที่ใช้ในการเปรียบเทียบคุณภาพของการเชื่อมโยงคะแนน

ตอนที่ 4 การตรวจสอบความเป็นพหุมิติ และความถูกต้องของโปรแกรมที่ใช้ในการจำลองข้อมูล

รายละเอียดของแต่ละตอนมีดังต่อไปนี้

ตอนที่ 1 เงื่อนไขที่ใช้ในการจำลองข้อมูล (Simulation Factor)

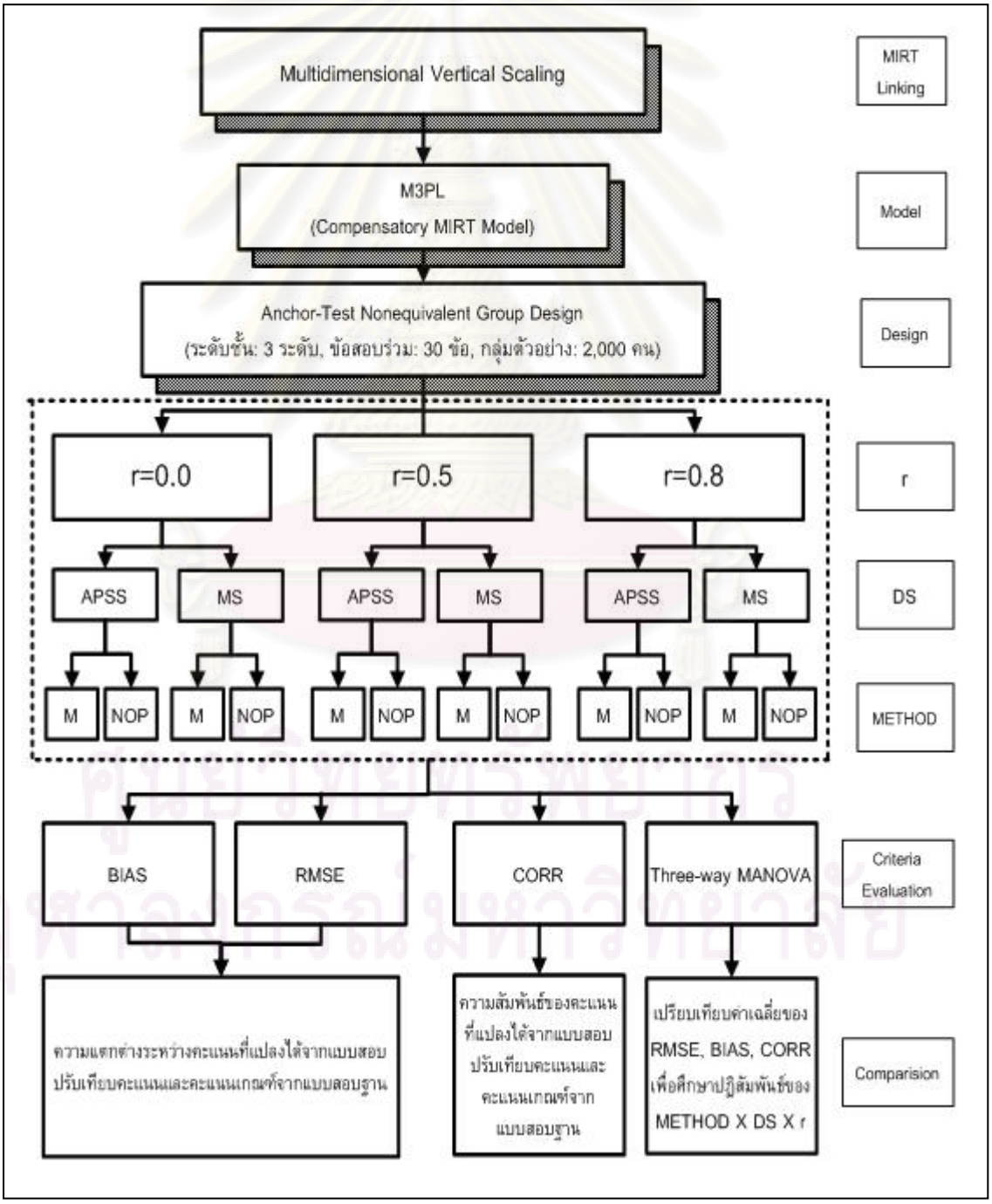
สิ่งสำคัญประการหนึ่งของการศึกษาโดยการจำลองข้อมูล คือสามารถทราบค่าพารามิเตอร์ที่แท้จริง (True Parameter) ทำให้สามารถเปรียบเทียบค่าพารามิเตอร์ที่แท้จริงกับค่าที่ได้จากการแปลงคะแนนจากแบบสอบปรับเทียบคะแนนไปยังแบบสอบฐาน การศึกษาโดยใช้การจำลองข้อมูลจึงมีประโยชน์ในการประเมินวิธีการเชื่อมโยงคะแนนทั้ง 2 วิธี เพื่อเปรียบเทียบผลและความคลาดเคลื่อนที่เกิดขึ้นจากวิธีการเชื่อมโยงคะแนนที่ต่างกันในสถานการณ์ที่ทำการศึกษา (Davey, Nering & Thompson, 1997; Bolt, 1999)

สำหรับการศึกษาในครั้งนี้ ทำการศึกษา 3 องค์ประกอบที่สำคัญ คือ *1) วิธีการเชื่อมโยงคะแนน (Linking Method)* โดยจำแนกเป็น 2 วิธี วิธีแรกเรียกว่าวิธี “M” (Min, 2003) ซึ่งเป็นการหมุนแกนแบบตั้งฉาก (Orthogonal Procrustes Method) ส่วนวิธีที่สองเรียกว่า วิธี “NOP” (Reckase & Martineau, 2004) เป็นวิธีการหมุนแกนแบบไม่ตั้งฉาก (Non-Orthogonal Procrustes Method) *2) โครงสร้างมิติความสามารถ (Dimensional Structure)* ของแบบสอบ มีสองลักษณะ ลักษณะแรกเป็นโครงสร้างที่ไม่ซับซ้อน (Approximate Simple Structure; APSS) ส่วนลักษณะที่สองเป็นโครงสร้างที่ซับซ้อนมากขึ้น (Mixed Structure; MS) และ 3) ระดับความสัมพันธ์ของมิติความสามารถ มี 3 ระดับ คือ ไม่มีความสัมพันธ์ ความสัมพันธ์ระดับกลาง และความสัมพันธ์ระดับสูง

จากการศึกษาที่ผ่านมาพบว่ากลุ่มผู้สอบประมาณ 2,000 คนขึ้นไปเหมาะสมสำหรับการศึกษาโดยโมเดล MIRT (Ackerman, 1994; Reckase, 1995; Yon, 2006) ดังนั้นในการศึกษาครั้งนี้จึงประเมินความคงที่ (Stability) ซึ่งเป็นคุณภาพของการเชื่อมโยงคะแนนภายใต้เงื่อนไขที่ศึกษา โดยใช้กลุ่มตัวอย่างระดับชั้นละ 2,000 คน จากทั้งหมด 3 ระดับชั้น คือระดับชั้นที่ 1-3 ออกแบบการเก็บรวบรวมข้อมูลโดยใช้รูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้แบบสอบร่วม (Anchor-Test Nonequivalent Group Design) ซึ่งมีข้อสอบร่วม (Common Items) แม้ว่าจากการศึกษาที่ผ่านมายังไม่มีผลการศึกษาที่แน่ชัดว่าจำนวนข้อสอบร่วมที่เหมาะสมในการเชื่อมโยงคะแนนโดยโมเดล MIRT เป็นเท่าใด แต่จากการศึกษาโดยส่วนใหญ่นิยมใช้จำนวนข้อสอบประมาณ 20% ของข้อสอบทั้งหมด ตามการศึกษาจำนวนข้อสอบร่วมที่เหมาะสมของ Angoff (1971) & Budescu (1985) ซึ่งเป็นเกณฑ์ที่ได้รับการยอมรับในปัจจุบันเพื่อใช้ในการศึกษาโมเดล UIRT และ MIRT (อ้างใน Min, 2003; Yon, 2006) เนื่องจากผู้วิจัยได้ทำการศึกษาเบื้องต้น (Pilot Study) กับจำนวนข้อสอบร่วมเพียง 20 ข้อ พบว่ามีค่าความเที่ยงของข้อสอบต่ำโดยเฉพาะในเงื่อนไขที่ไม่มีความสัมพันธ์ระหว่างมิติความสามารถ (Corr. = 0.0) ซึ่งมีความเที่ยงอยู่ในช่วง 0.40-0.55 แต่เมื่อความสัมพันธ์ระหว่างมิติสูงขึ้นค่าความเที่ยงมีค่าเพิ่มขึ้นแต่มีค่าไม่เกิน 0.70 สาเหตุสำคัญคือข้อสอบมีน้อยเกินไปสำหรับใช้ในการศึกษา 2 มิติในกลุ่มตัวอย่างขนาดใหญ่ โดยผลการศึกษาดังกล่าวสอดคล้องกับการศึกษาของ Li & Lissize (2000) และ Reckase & Matineua (2000) ที่พบว่าจำนวนข้อสอบร่วมเป็นสิ่งที่สำคัญ โดย

จำนวนข้อสอบร่วมที่ยาวกว่าย่อมให้ผลที่ดีกว่า ซึ่งจากการศึกษาของ Li & Lissitz (2000) ได้ศึกษากับจำนวนของข้อสอบร่วม 15 ข้อ และ 25 ข้อ ส่วน Reckase & Martineau (2004, 2006) ได้ใช้ข้อสอบร่วม ระดับละ 25 ข้อ ข้อค้นพบจากการศึกษาดังกล่าวพบว่าข้อสอบร่วมภายในน้อยเกินไปกับกลุ่มผู้สอบที่มีขนาดใหญ่ ดังนั้นการวิจัยในครั้งนี้ผู้วิจัยจึงเพิ่มจำนวนข้อสอบร่วมจาก 20 ข้อ เป็น 30 ข้อ ซึ่งคิดเป็น 30 % ของข้อสอบทั้งหมด เนื่องจากมีความเป็นตัวแทนของเนื้อหาทั้งในมิติที่ 1 และมิติที่ 2 มากกว่า

จากองค์ประกอบหรือเงื่อนไขที่กล่าวมาข้างต้นสามารถเสนอด้วยแบบแผนการจำลองข้อมูลดังนี้



แผนภาพที่ 3.1 แบบแผนการจำลองข้อมูล

จากแบบแผนการจำลองข้อมูลดังกล่าวข้างต้น มีรายละเอียดของเงื่อนไขที่ใช้ในการจำลองข้อมูล ดังต่อไปนี้

1. วิธีการเชื่อมโยงคะแนน (Linking Method)

วิธีการเชื่อมโยงคะแนนทั้งสองวิธี คือ M และ NOP ผู้วิจัยจะทำการเปรียบเทียบความถูกต้องของการแปลงคะแนนจากแบบสอบปรับเทียบคะแนนไปยังแบบสอบฐาน โดยความถูกต้องของการเชื่อมโยงคะแนนจะประเมินจากการเปรียบเทียบค่าพารามิเตอร์ของข้อสอบที่ประมาณค่าได้จากแบบสอบปรับเทียบคะแนน (แบบสอบของระดับชั้นที่ต่ำกว่า) ไปยังแบบสอบฐาน (แบบสอบในระดับชั้นที่สูงกว่า) ภายหลังการเชื่อมโยงคะแนนให้อยู่ในสเกลที่สามารถเปรียบเทียบกันได้ โดยมีรายละเอียดของแต่ละวิธี ดังต่อไปนี้

1.1 การเชื่อมโยงคะแนนที่มีการหมุนแกนแบบตั้งฉากโดยวิธี M (M method)

วิธี M เป็นวิธีการเชื่อมโยงคะแนนที่พัฒนาโดย Min (2003) ซึ่งทำการพิจารณาค่าพารามิเตอร์การเลื่อนแกนที่เป็นสเกลลา (Scalar Dilation Parameter) ตามวิธีของ Li & Lissitz (2000) แล้วได้พบจุดอ่อนของวิธีการดังกล่าวว่าไม่เพียงพอสำหรับการหดหรือการขยายสเกลการเชื่อมโยงคะแนนที่มีลักษณะหลายมิติ ซึ่งค่าพารามิเตอร์การเลื่อนแกน (Scalar Dilation Parameter) สามารถขยายสเกลไปยังมิติแต่ละมิติให้อยู่ในลักษณะที่เป็นรูปแบบเหมือนกันโดยตรง แต่เมื่อแยกการประมาณค่าตามโมเดล MIRT ที่มีหลากหลายรูปแบบอาจส่งผลให้สเกลขยายไปยังมิติในระดับต่างกัน ที่อาจมีปัญหาในการประมาณค่าสำหรับการปรับเทียบสเกลแนวตั้งที่มีหลายระดับต่างกัน ดังนั้น Min จึงทำการพัฒนาวิธีการดังกล่าวโดยมุ่งที่จะเชื่อมโยงคะแนนในลักษณะที่เป็นขั้นตอนสมการทางคณิตศาสตร์ (Algorithm) โดยมีชื่อเรียกว่า “M” ซึ่งทำการแทนค่าสเกลลาการเลื่อนแกนด้วยเมตริกซ์ (Diagonal Dilation Matrix) ทำให้ผลที่ได้มีประสิทธิภาพมากขึ้น โดยวิธีการของ Min (2003) แตกต่างจากวิธีของ Li & Lissitz (2000) คือ ยอมให้มีการเปลี่ยนแปลงหน่วยความยาวในแต่ละมิติมากกว่าการเปลี่ยนแปลงค่าสเกลลาในมิติทั้งหมดโดยภาพรวม และแตกต่างจากการศึกษาตามวิธี NOP (Non-Orthogonal Procrustes) (Reckase & Martineau, 2004) โดยวิธี M จะยึดการหมุนแกน และระยะของมาตรวัด อีกทั้งมีขั้นตอนที่แตกต่างกัน โดยใช้การหมุนแกนแบบตั้งฉาก (Orthogonal Procrustes Rotation) วิธีการดังกล่าวสามารถเขียนสมการในการแปลงคะแนน ดังต่อไปนี้

$$a_i^* = a_i TK \quad \text{.....(3.1)}$$

$$d_i^* = d_i - a_i^* Tm \quad \text{.....(3.2)}$$

$$\theta_j^* = K^{-1}\left(T^{-1}\theta_j + m\right) \quad \text{.....(3.3)}$$

เมื่อ K_{ij} เป็น "diagonal dilation matrix”

T แทน เมตริกซ์การหมุนแกนแบบมุมฉาก (Orthogonal Rotation Matrix) สำหรับใช้ในการกำหนดทิศทาง

M แทน เวคเตอร์การแปลงคะแนนสำหรับการกำหนดตำแหน่ง (Location)

a_i แทน เวคเตอร์ของค่าพารามิเตอร์ที่มีความสัมพันธ์กับอำนาจจำแนกของข้อสอบข้อที่ i ในแต่ละมิติที่มีจำนวน m มิติ

D แทน ค่าพารามิเตอร์ที่มีลักษณะเป็นค่าสเกลลาที่มีความสัมพันธ์กับค่าความยากของข้อสอบข้อที่ i

θ_j แทน เวคเตอร์ของค่าพารามิเตอร์ที่อธิบายความสามารถของผู้สอบคนที่ j ในมิติที่ m (m-Dimension Space)

หมายเหตุ: * เป็นสัญลักษณ์ที่บ่งชี้ถึงค่าที่มีการแปลงคะแนนจากแบบสอบปรับเทียบคะแนนไปยังแบบสอบฐาน

จากสมการดังกล่าวจะเห็นได้ว่า K เป็นเมตริกซ์ของค่าพารามิเตอร์การเลื่อนแกน (Diagonal Dilation Matrix) ซึ่งสถานการณ์ที่ทำการศึกษาในครั้งนี้ ระบุให้ $K = \begin{bmatrix} k_1 & 0 \\ 0 & k_2 \end{bmatrix}$ เมื่อ k_1 ในเมตริกซ์ K แทน ส่วนประกอบของค่าการเลื่อนแกน (Dilation) ในมิติที่ 1 ทำนองเดียวกัน k_2 ในเมตริกซ์ K แทน ส่วนประกอบของ ค่าการเลื่อนแกน (Dilation) ในมิติที่ 2 สำหรับส่วนประกอบอื่นที่นอกเหนือแนวทแยงกำหนดให้เป็น 0 เนื่องจากมาตรวัดจะถูกนำไปแทนที่ยังแกนของมิติที่ต้องการวัดมากกว่าส่วนที่ไม่ใช่มุมฉาก ในแต่ละคอลัมน์ของ K จะประกอบด้วยจำนวนที่มีค่าไม่เท่ากับ 0 เพื่อให้สอดคล้องกับมิติที่ต้องการศึกษาในแต่ละแกน

1.2 การเชื่อมโยงคะแนนที่มีการหมุนแกนแบบไม่ตั้งฉากโดยวิธี NOP (Non-Orthogonal Procrustes Method)

Reckase & Martineau (2004) ได้เสนอแนะวิธีการแปลงคะแนนแบบไม่ตั้งฉาก ซึ่งเป็นวิธีการแปลงคะแนนจากมิติของแบบสอบปรับเทียบคะแนนไปยังมิติของแบบสอบฐาน โดยขึ้นอยู่กับความสอดคล้องที่เหมาะสมมากที่สุดของมิติในแบบสอบปรับเทียบคะแนนไปยังแบบสอบฐาน ข้อได้เปรียบของวิธี NOP คือ มีการตัดเมตริกซ์การเลื่อนแกน (Dilation Matrix) ซึ่งมีค่าเท่ากับ K ออกจากสมการ รวมถึงการเลื่อนแกนที่เป็นสเกลลา (Dilation Scalar) ด้วยเช่นกัน

จากการศึกษาของ Mulaik (1972) ที่ใช้กระบวนการหมุนแกนแบบมุมแหลม (Oblique Procrustes) สามารถแสดงสมการเมตริกซ์การหมุนแกน (Rotation Matrix) ดังต่อไปนี้

$$T = (A'A)^{-1}A'B \quad \text{.....(3.4)}$$

$$a_i^* = a_i T \quad \text{.....(3.5)}$$

$$d_i^* = d_i + a_i^* Tm \quad \text{.....(3.6)}$$

$$\theta_j^* = \left(T^{-1}\theta_j - m\right) \quad \text{.....(3.7)}$$

เมื่อ T แทน เมตริกซ์การหมุนแกน (Rotation Matrix) สำหรับการบอกตำแหน่ง

A แทน เมตริกซ์ของค่าพารามิเตอร์อำนาจจำแนกของแบบสอบปรับเทียบคะแนน

B แทน เมตริกซ์ของค่าพารามิเตอร์อำนาจจำแนกของแบบสอบฐาน

M แทน เวคเตอร์การแปลงคะแนนสำหรับการกำหนดตำแหน่ง (Location)

a_i แทน เวคเตอร์ของค่าอำนาจจำแนกของข้อสอบข้อที่ i ในแต่ละมิติที่มีจำนวน m มิติ

D แทน ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i

θ_j แทน เวคเตอร์ของค่าพารามิเตอร์ความสามารถของผู้สอบคนที่ j ในมิติที่ m (m-Dimension Space)

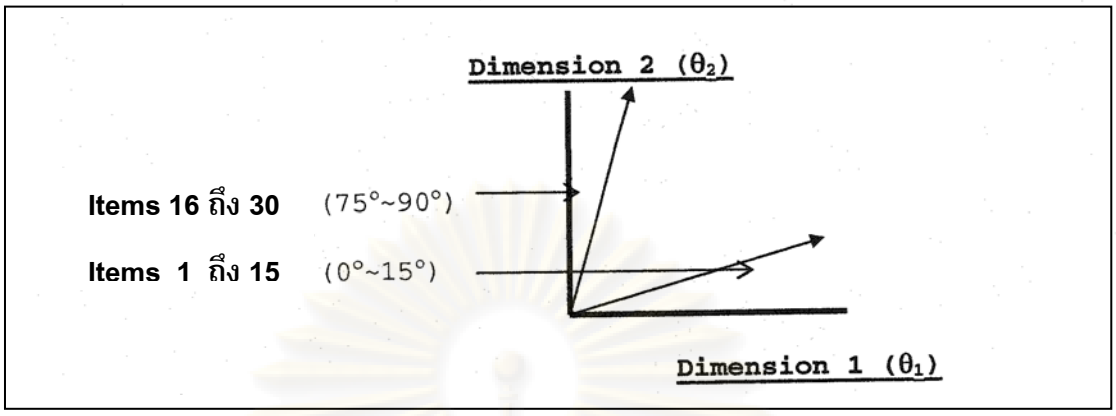
หมายเหตุ: * เป็นค่าพารามิเตอร์ของการแปลงคะแนน (Transform Parameter) ไปยังแบบสอบฐาน

2. โครงสร้างมิติความสามารถ (Dimensional Structure)

โดยทั่วไปโครงสร้างมิติความสามารถจะประกอบไปด้วย 3 ลักษณะ คือ โครงสร้างอย่างง่าย (Simple Structure; SS) โครงสร้างที่ไม่ซับซ้อน (Approximate Simple Structure; APSS) และ โครงสร้างที่ซับซ้อนมากขึ้น (Mixed Structure; MS) แต่การศึกษาครั้งนี้เลือกศึกษาโครงสร้างมิติความสามารถเพียง 2 ลักษณะ คือ โครงสร้างที่ไม่ซับซ้อน และโครงสร้างที่ซับซ้อนมากขึ้น เนื่องจากสอดคล้องกับสถานการณ์จริงมากกว่าโครงสร้างอย่างง่าย (Roussos, Stout, & Marden, 1998; Min, 2003; Yon, 2006) ซึ่งมีรายละเอียดของของแต่ละลักษณะที่ทำการศึกษา ดังนี้

2.1 โครงสร้างที่ไม่ซับซ้อน (Approximate Simple Structure; APSS)

โครงสร้างที่ไม่ซับซ้อน (APSS) มีลักษณะโครงสร้างที่แต่ละข้อของแบบสอบมีความสัมพันธ์กันสูงในมิติที่หนึ่งมากกว่ามิติที่สอง นั่นคือ กลุ่มของข้อสอบจำนวนหนึ่งมีค่าอำนาจจำแนกที่สูงในมิติเดียวกันแต่มีค่าอำนาจจำแนกต่ำสำหรับอีกมิติหนึ่ง ในการศึกษาครั้งนี้ผู้วิจัยสนใจศึกษาแบบสอบที่มีลักษณะ 2 มิติ โดยมีข้อสอบร่วมในแต่ละระดับทั้งหมด 20 ข้อ โดยในแต่ละมิติมีจำนวนข้อสอบเท่ากันมิติละ 15 ข้อ ลักษณะเวคเตอร์ของข้อสอบในมิติที่ 1 (Θ_1) จะทำมุมกับแกนของมิติที่ 1 ตั้งแต่ 0°–15° ($0^\circ \leq \alpha_k \leq -15^\circ$) เมื่อ α_k เป็นมุมระหว่างเวคเตอร์ในมิติที่ 1 กับแกนของมิติที่ 1 ในขณะที่ลักษณะเวคเตอร์ของข้อสอบในมิติที่ 2 (Θ_2) จะทำมุมกับแกนของมิติที่ 1 ช่วงมุมตั้งแต่ 75°–90° ($90-\alpha_k$) ซึ่งโครงสร้างดังกล่าวสามารถแสดงได้ดังแผนภาพที่ 3.2



แผนภาพที่ 3.2 แบบสอบ 2 มิติที่มีลักษณะโครงสร้างที่ไม่ซับซ้อน
(Approximate Simple Structure; APSS)

จากโครงสร้างที่ไม่ซับซ้อนดังแผนภาพ สามารถแสดงตารางจำนวนชุดข้อสอบย่อยที่ศึกษาเพื่อนำมาปรับเทียบสเกลแนวตั้งของนักเรียน 2 ระดับชั้นที่อยู่ติดกัน โดยการศึกษาครั้งนี้จะมีการทำซ้ำ (Replication) ทั้งหมด 20 ครั้ง ซึ่งในความหมายของโครงสร้างที่ไม่ซับซ้อน จะทำการสร้างแบบสอบ 2 ชุด แต่ละชุดมีข้อสอบ 15 ข้อ ข้อสอบชุดแรกมีลักษณะอยู่ในมิติที่ 1 มากกว่ามิติที่ 2 ในขณะที่ข้อสอบในชุดที่ 2 ใกล้เคียงกับมิติที่ 2 มากกว่ามิติที่ 1 ดังแสดงในตารางที่ 3.1

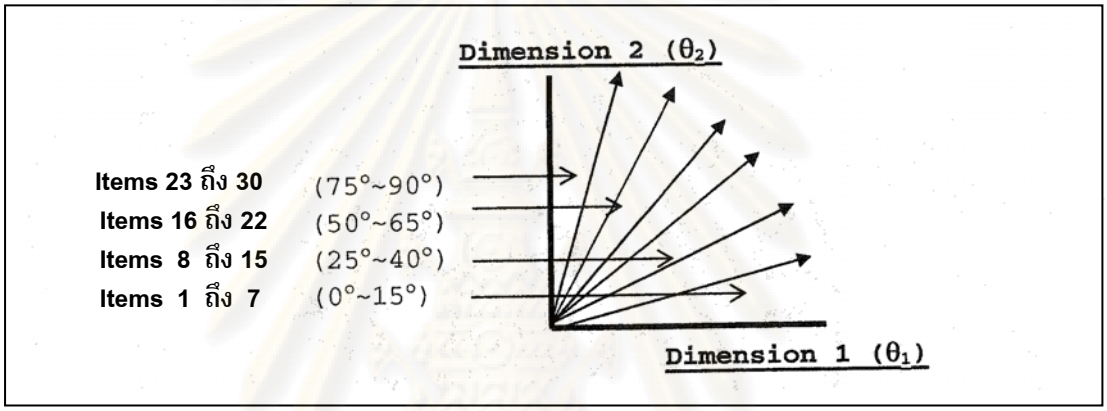
ตารางที่ 3.1 จำนวนของข้อสอบต่อแบบสอบแต่ละชุดและขนาดของมุมในโครงสร้าง APSS

ลำดับข้อสอบ	แบบสอบในระดับชั้นที่ต่ำกว่า (lower grade level)		ลำดับข้อสอบ	แบบสอบในระดับชั้นที่สูงกว่า (higher grade level)	
	ขนาดของมุมจาก (Angles From)			ขนาดของมุมจาก (Angles From)	
	มิติที่ 1 (θ_1)	มิติที่ 2 (θ_2)		มิติที่ 1 (θ_1)	มิติที่ 2 (θ_2)
1-15	0°–15°	75°–90°	1-15	0°–15°	75°–90°
16-30	75°–90°	0°–15°	16--30	75°–90°	0°–15°

2.2 โครงสร้างที่ซับซ้อนมากขึ้น (Mixed Structure; MS)

โครงสร้างที่ซับซ้อนมากขึ้น (MS) เป็นลักษณะโครงสร้างที่สอดคล้องกับสถานการณ์จริงมากกว่าโครงสร้างอย่างง่าย (SS) เนื่องจากโครงสร้างโดยทั่วไปไม่ได้มีเพียงมิติใดมิติหนึ่งเท่านั้น แต่ประกอบด้วยหลายๆ มิติ ผู้วิจัยจึงเลือกที่จะศึกษา ลักษณะโครงสร้างที่ซับซ้อนมากขึ้น จะมีการวัดในแต่ละมิติและวัดมิติโดยรวม (Composites of Dimensions) (Kim, 1994) ซึ่งในการศึกษาครั้งนี้จะศึกษาเช่นเดียวกับโครงสร้างอย่างง่าย แต่ขนาดของมุมระหว่างข้อสอบกับแกนของมิติจะแตกต่างกันไป โดยมีแบบสอบทั้งหมด 4 ชุดย่อย แบบสอบย่อย 2 ชุดแรก ชุดละ 7 ข้อ จะแยกเป็นมิติที่ 1 (0°–15°) ประกอบไปด้วยข้อที่ 1-7 และ มิติที่ 2 (75°–90°) ซึ่ง

ประกอบไปด้วยข้อที่ 23-30 ส่วนแบบสอบย่อยอีก 2 ส่วนที่เหลือ คือ ข้อที่ 8-15 และ 16-22 จะวัดในลักษณะของมิติรวม ข้อที่ 8-15 ข้อสอบมีลักษณะใกล้เคียงกับมิติที่ 1 มากกว่ามิติที่ 2 ในขณะที่ข้อ 16-22 ข้อสอบมีลักษณะใกล้เคียงกับมิติที่ 2 มากกว่ามิติที่ 1 โดยการศึกษาครั้งนี้ได้นำแนวคิดของ Wang (1986) ที่มีการนำเสนอแนวคิดแกนโดยรวม (Reference Composite) ในลักษณะของการศึกษาแบบพหุมิติ (Multivariate Space) ซึ่งค่าที่ได้สามารถสะท้อนถึงค่าอำนาจจำแนกรวม (MDISCs) และค่าความแปรปรวนความแปรปรวนร่วมของการกระจายความสามารถที่มีลักษณะแบบหลายมิติ (Ackerman, 1992) โครงสร้างดังกล่าวสามารถแสดงได้ดังแผนภาพที่ 3.3



แผนภาพที่ 3.3 แบบสอบ 2 มิติที่มีลักษณะโครงสร้างที่ซับซ้อนมากขึ้น (Mixed Structure; MS)

จากโครงสร้างมิติความสามารถที่ซับซ้อนมากขึ้น สามารถแสดงตารางของแบบสอบย่อยทั้งหมด 4 ชุดย่อย มี 1 ชุดใกล้เคียงกับมิติที่ 1 อีก 1 ชุด ใกล้เคียงกับมิติที่ 2 ส่วนแบบสอบอีก 2 ชุดที่เหลือ จะศึกษาในลักษณะของมิติโดยรวม ดังแสดงในตารางที่ 3.2

ตารางที่ 3.2 จำนวนของข้อสอบต่อแบบสอบแต่ละชุดและขนาดของมุมในโครงสร้าง MS

ลำดับข้อสอบ	แบบสอบในระดับชั้นที่ต่ำกว่า (lower grade level)		ลำดับข้อสอบ	แบบสอบในระดับชั้นที่สูงกว่า (higher grade level)	
	ขนาดของมุมจาก (Angles From)			ขนาดของมุมจาก (Angles From)	
	มิติที่ 1 (θ_1)	มิติที่ 2 (θ_2)		มิติที่ 1 (θ_1)	มิติที่ 2 (θ_2)
1-7	0°–15°	75°–90°	1-7	0°–15°	75°–90°
8-15	25°–40°	50°–65°	8-15	25°–40°	50°–65°
	Composite trait 1>2			Composite trait 1>2	
16-22	50°–65°	25°–40°	16-22	50°–65°	25°–40°
	Composite trait 2>1			Composite trait 2>1	
23-30	75°–90°	0°–15°	23-30	75°–90°	0°–15°

3. ระดับความสัมพันธ์ของมิติความสามารถ (Correlation Levels between Traits)

ความสัมพันธ์ระหว่างมิติความสามารถ เป็นค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าการแจกแจงความสามารถของผู้สอบในมิติที่ 1 กับค่าการแจกแจงความสามารถของผู้สอบในมิติที่ 2 ถ้าค่าสัมประสิทธิ์สหสัมพันธ์มีค่าเท่ากับ 0 แสดงว่าการประมาณค่าความสามารถของผู้สอบในมิติที่ 1 กับมิติที่ 2 เป็นอิสระต่อกัน แต่ถ้าค่าสัมประสิทธิ์สหสัมพันธ์มีค่าไม่เท่ากับ 0 แสดงว่าการประมาณค่าความสามารถของผู้สอบในมิติที่ 1 กับมิติที่ 2 มีความเกี่ยวข้องสัมพันธ์กัน โดยงานวิจัยส่วนใหญ่ที่ทำการศึกษา จะมีการกำหนดให้ไม่มีความสัมพันธ์ระหว่างมิติความสามารถ (ρ =0) นั่นคือ ระดับความสามารถของผู้สอบในอีกมิติหนึ่งไม่มีความสัมพันธ์กับความสามารถในอีกมิติหนึ่ง ซึ่งผู้วิจัยมีความเห็นว่าอาจไม่สอดคล้องกับสภาพความเป็นจริงในบางสถานการณ์ที่ทำการศึกษา ที่ระดับความสามารถของผู้สอบในอีกมิติหนึ่งมีความสัมพันธ์กับความสามารถในอีกมิติ จะมากหรือน้อยย่อมขึ้นอยู่กับเนื้อหาที่ทำการศึกษาเป็นสำคัญ โดยเฉพาะอย่างยิ่ง แบบสอบวัดผลสัมฤทธิ์ทางการเรียนที่มีการบูรณาการข้ามเนื้อหาสาระอย่างเห็นได้ชัด (Torre & Patz, 2006)

จากการศึกษาของ CTB/McGraw-Hill (2002) และ Johnson & Carlson (1994) พบว่า แบบสอบวัดผลสัมฤทธิ์ทางการเรียนจะมีความสัมพันธ์ระหว่างมิติความสามารถอยู่ในระดับสูง มีค่าตั้งแต่ 0.7 ขึ้นไป โดยเมื่อนำมาหาค่าเฉลี่ยของความสัมพันธ์ระหว่างมิติความสามารถพบว่า อยู่ในระดับ 0.8 ซึ่งทำการศึกษาจากแบบสอบวัดผลสัมฤทธิ์ทางการเรียนเพื่อวัดความก้าวหน้าระดับชาติของประเทศสหรัฐอเมริกา (National Assessment of Education Progress; NAEP) ตั้งแต่ปี 1992-2007 เมื่อผู้วิจัยทำการวิเคราะห์กับหลักสูตรทางการศึกษาในประเทศไทยโดยเฉพาะในกลุ่มสาระการเรียนรู้คณิตศาสตร์ ที่มีการแบ่งเนื้อหาออกเป็น 5 กลุ่ม และมีการให้น้ำหนักความสำคัญในแต่ละกลุ่มเนื้อหาที่ใกล้เคียงกัน โดยมีความสัมพันธ์อยู่ในระดับสูงระหว่างมิติความสามารถเช่นเดียวกัน ดังนั้นผู้วิจัยจึงสนใจศึกษากับความสัมพันธ์ระหว่างมิติความสามารถในระดับสูงด้วย ซึ่งกำหนดให้มีค่าเท่ากับ 0.8 (r=0.8) เนื่องจากเป็นค่าที่มีความเป็นตัวแทนที่ได้จากการหาค่าเฉลี่ยของระดับความสัมพันธ์ที่มีค่าอยู่ในระดับสูง นอกจากนี้เพื่อให้การศึกษาในครั้งนี้มีความครอบคลุมมากขึ้นผู้วิจัยจึงสนใจศึกษากับความสัมพันธ์ระหว่างมิติความสามารถที่อยู่ในระดับกลางด้วย ซึ่งกำหนดให้มีค่าเท่ากับ 0.5 (r=0.5) (Min, 2003)

ดังนั้นในการศึกษาครั้งนี้ ผู้วิจัยสนใจศึกษาความสัมพันธ์ 3 ระดับ ประกอบด้วย ไม่มีความสัมพันธ์ (r=0.0) มีความสัมพันธ์ระดับกลาง (r=0.5) และมีความสัมพันธ์ระดับสูง (r=0.8)

ตอนที่ 2 การศึกษาการจำลองข้อมูล (Simulation Study)

การวิจัยในครั้งนี้ เงื่อนไขที่ใช้ในการศึกษาประกอบด้วย วิธีการที่ใช้ในการเชื่อมโยงคะแนน (Linking Method) จำนวน 2 วิธี โครงสร้างมิติความสามารถ (Dimensional Structure) จำนวน 2 ลักษณะ และระดับความสัมพันธ์ของมิติความสามารถ 3 ระดับ ทำให้มีเงื่อนไขที่ทำการศึกษาทั้งหมด 12 เงื่อนไข (2x2X3) แม้ว่าการศึกษาที่ผ่าน ยังไม่มีผลงานวิจัยที่ชัดเจนว่าควรมีการสร้างการจำลองข้อมูลโดยการทำซ้ำ (Replication) ทั้งหมดกี่ครั้ง เพื่อให้ผลที่เกิดขึ้นมี

ความเที่ยงมากที่สุด แต่อย่างไรก็ตามจากการศึกษาของ Harwell, Hsu & Kirisci (1996) ที่พบว่าในการศึกษาโดยใช้โมเดล IRT เป็นฐานควรมีการทำซ้ำอย่างน้อย 20 ครั้ง ดังนั้นการศึกษาในครั้งนี้จึงมีการสร้างค่าพารามิเตอร์ที่แท้จริงของข้อสอบและแบบแผนการตอบของนักเรียนแต่ละระดับโดยการทำซ้ำทั้งหมด 20 ครั้ง เพื่อให้การประมาณค่ามีความเที่ยงเพิ่มขึ้น ซึ่งในขั้นตอนของการจำลองข้อมูลประกอบไปด้วย 3 ขั้นตอน คือ

1. การสร้างค่าพารามิเตอร์ที่แท้จริงของข้อสอบและแบบแผนการตอบของนักเรียนแต่ละระดับ
2. การประมาณค่าพารามิเตอร์ตามโมเดล MIRT
3. การเชื่อมโยงคะแนนหรือการปรับเทียบคะแนน

โดยในแต่ละขั้นตอนมีรายละเอียด ดังนี้

2.1 การสร้างค่าพารามิเตอร์ที่แท้จริงของข้อสอบและแบบแผนการตอบของนักเรียนแต่ละระดับ (Generation of True Item Parameters and Response Patterns for Each Grade Level)

ในการสร้างค่าพารามิเตอร์ที่แท้จริง ค่าอำนาจจำแนกรวม (MDISC) และ ค่าความยากรวม (MDIFF) นั้นจะต้องเป็นค่าที่สอดคล้องกับความเป็นจริง ซึ่งเป็นช่วงของค่าอำนาจจำแนกรวม (MDISC) และ ค่าความยากรวม (MDIFF) ถูกกำหนดขึ้นจากค่าที่ได้จากแบบสอบวัดผลสัมฤทธิ์ทางการเรียนที่เกิดขึ้นในสภาพจริง โดยสอดคล้องกับการศึกษาของ Roussos et al. (1998), Min (2003) & Yon (2006) โดยในที่นี้ ผู้วิจัยเลือกใช้ค่าที่นำเสนอโดย Yon (2006) เนื่องจากเป็นค่าที่สอดคล้องกับการศึกษาในครั้งนี้ โดยมีการแบ่งเป็นค่าที่ได้จากแบบสอบในระดับที่ต่ำกว่ากับค่าที่ได้จากแบบสอบในระดับที่สูงกว่า

อย่างไรก็ตามในการเลือกค่าที่นำมาใช้ศึกษาอาจจะไม่ครอบคลุมกับระดับหรือช่วงค่าของความยากทั้งหมด ซึ่งค่าที่ได้เป็นค่าที่คาดหวังว่าจะเกิดขึ้นจริง โดยศึกษาเฉพาะค่าที่มีความสัมพันธ์กับข้อสอบที่ยาก (Difficult Items) ซึ่งขัดแย้งกับการศึกษาของ Roussos และคณะ (1998) ที่ทำการศึกษากับข้อสอบที่ง่าย (Easy Items) ด้วย อย่างไรก็ตามการศึกษาก่อนหน้านี้ไม่ได้ศึกษาในลักษณะของการปรับเทียบสเกลแนวตั้ง (Vertical Scaling) นอกจากนี้มีการประมาณค่าพารามิเตอร์ความยากและค่าพารามิเตอร์อำนาจจำแนก เพียง 1 ระดับชั้น ดังนั้นการวิเคราะห์ในครั้งนี้ได้จำลองข้อมูลโดยยึดข้อมูลหลายระดับชั้นในแต่ละชุดของแบบสอบ โดยในเงื่อนไขนี้จะใช้คอมพิวเตอร์ในการคำนวณ เพื่อใช้ในการประมาณค่าพารามิเตอร์ ซึ่งข้อสอบใดมีค่าความยากของข้อสอบเกินขอบเขตมากเกินไป จะไม่นำมาศึกษาและคัดเลือกออกจากการวิเคราะห์ ซึ่งจำเป็นที่จะต้องทำการศึกษาในช่วงที่ไม่มากจนเกินไป ค่าที่ได้ต้องเป็นผลมาจากการสุ่มรวมถึงคัดเลือกข้อที่ง่ายมาศึกษาด้วยเช่นกัน การรวมค่าอำนาจจำแนกรวมและค่าความยากรวมมาศึกษาพร้อมกันในครั้งนี้จะสอดคล้องกับการประมาณค่าของข้อสอบที่มีความยากได้

เหมาะสมกว่าวิธีการที่ศึกษาโดย Roussios et al. (1998) โดยชุดของข้อสอบถูกแบ่งออกเป็น 5 ระดับ ในแต่ระดับจะมีค่าความยากที่แตกต่างกัน ซึ่งมีการกำหนดค่าที่เป็นไปได้ทั้งหมด ดังเสนอในตารางที่ 3.3 เพื่อใช้จำลองค่าพารามิเตอร์อำนาจจำแนก (a_1 และ a_2) และค่าความยากของข้อสอบ (d) สำหรับนักเรียนระดับที่ 2 ทำแบบสอบในระดับที่ 2 และนักเรียนในระดับที่ 3 ทำแบบสอบในระดับที่ 3 ซึ่งเหมาะสำหรับนักเรียนที่อยู่ในระดับที่สูงกว่า เนื่องจากมีค่าความยากกง่ายรวมน้อย และตารางที่ 3.4 ใช้จำลองค่าพารามิเตอร์อำนาจจำแนก (a_1 และ a_2) และค่าความยากของข้อสอบ (d) สำหรับนักเรียนระดับที่ 1 ทำแบบสอบในระดับที่ 2 และนักเรียนในระดับที่ 2 ทำแบบสอบในระดับที่ 3 ซึ่งเหมาะสำหรับนักเรียนที่อยู่ในระดับที่ต่ำกว่า เนื่องจากมีค่าความยากกง่ายรวมมากกว่า

ตารางที่ 3.3 ลักษณะการกระจายค่าความยากรวมและค่าอำนาจจำแนกรวมในแต่ละกลุ่มที่แตกต่างกันเพื่อใช้จำลองค่าพารามิเตอร์อำนาจจำแนก (a_1 และ a_2) และค่าความยากของข้อสอบ (d) สำหรับนักเรียนที่อยู่ในระดับที่สูงกว่า (Yon, 2006)

การแบ่งช่วงการกระจายของค่าพารามิเตอร์	ค่าอำนาจจำแนกรวม (MDISC)	ค่าความยากรวม (MDIFF)
1	0.6	-0.5
2	0.8	2.0
3	1.0	1.0
4	1.2	0.0
5	1.3	2.2
ค่าเฉลี่ย	**1.0**	**0.9**

ตารางที่ 3.4 ลักษณะการกระจายค่าความยากรวมและค่าอำนาจจำแนกรวมในแต่ละกลุ่มที่แตกต่างกันเพื่อใช้จำลองค่าพารามิเตอร์อำนาจจำแนก (a_1 และ a_2) และค่าความยากของข้อสอบ (d) สำหรับนักเรียนที่อยู่ในระดับที่ต่ำกว่า (Yon, 2006)

การแบ่งช่วงการกระจายของค่าพารามิเตอร์	ค่าอำนาจจำแนกรวม (MDISC)	ค่าความยากรวม (MDIFF)
1	0.6	0.0
2	0.8	2.5
3	1.0	1.5
4	1.2	0.5
5	1.3	3.0
ค่าเฉลี่ย	**1.0**	**1.5**

จากโครงสร้างที่แสดงในตารางที่ 3.1 และ 3.2 ซึ่งเป็นโครงสร้างมิติความสามารถทั้ง 2 ลักษณะ คือโครงสร้างที่ไม่ซับซ้อนและโครงสร้างที่ซับซ้อนมากขึ้น โดยมุมระหว่างข้อสอบในแต่ละมิติ แทนด้วยสัญลักษณ์ α_{ik}° จะถูกสุ่มจากการกระจายแบบสม่ำเสมอ (Uniform Distribution) ดังนั้นการแสดงค่าน้ำหนักของข้อสอบ (Loading) และค่า MDISC ในตารางที่ 3.3 และ 3.4 ในแต่ละ ค่าที่มีขนาด α_{ik}° ซึ่งเป็นค่าที่แสดงน้ำหนักของข้อสอบข้อที่ i ในมิติที่ k สามารถคำนวณได้ ดังสมการที่ 3.8

$$MDISC_i = \sqrt{\sum_{k=1}^{m} \alpha_{ik}^2} \quad \text{.....(3.8)}$$

เมื่อ $MDISC_i$ แทน ค่าอำนาจจำแนกรวมของข้อสอบข้อที่ i ในมิติที่ m

M แทน จำนวนของมิติในมิติของ Θ (Θ-space)

a_{ik} แทน ส่วนประกอบของเวคเตอร์ที่ a_i *ในมิติที่* k

จากสมการ 3.8 จะเห็นว่า ค่าอำนาจจำแนกรวมของข้อสอบ (MDISC) เป็นสัดส่วนของฟังก์ชันการตอบข้อสอบในตำแหน่งที่มีค่าความชันสูงสุด (Steepest Slope) ซึ่งมีความคล้ายคลึงกับค่าอำนาจจำแนกของข้อสอบตามโมเดล UIRT (Carlson, 1987; Reckase & McKinly, 1991) โดยในการวิจัยวิจัยครั้งนี้ผู้วิจัยได้ทำการสุ่มมุมจากการกระจายแบบสม่ำเสมอ (Uniform Distribution) โดยอาศัยหลักการของการประมาณค่าพารามิเตอร์การแปลงคะแนนที่เรียกว่า “Transformation Parameter Estimates to a Common Coordinate System” ตามหลักการแปลงคะแนนของโมเดล MIRT ที่ว่า

“เมื่อทำการเลื่อนแกนการแปลงคะแนนจากระดับชั้นหนึ่งไปยังอีกระดับชั้นหนึ่ง ค่าที่ไม่เปลี่ยนแปลงคือค่าอำนาจจำแนกของข้อสอบ (a_1 และ a_2) แต่สิ่งที่เปลี่ยนแปลงไปคือค่าความสามารถของผู้สอบและค่าความยากของข้อสอบ (d)” (Reckase, 2007)

จากหลักการดังกล่าวผู้วิจัยจึงกำหนดให้ขนาดของมุมในแต่ละระดับมีค่าเท่ากัน เนื่องจากการกำหนดขนาดของมุมมีผลต่อค่าอำนาจจำแนกของข้อสอบโดยตรง ซึ่งการศึกษาครั้งนี้ถ้าอยู่ในจำนวนข้อเดียวกันและมีโครงสร้างเดียวกันจะมีขนาดมุมเท่ากันในทุกระดับตั้งแต่ระดับที่ 1-3 ซึ่งผลการสุ่มมุมจากการกระจายแบบสม่ำเสมอ (Uniform Distribution) โดยใช้โปรแกรม MATLAB ตามโครงสร้างที่ไม่ซับซ้อน (APSS) ดังตารางที่ 3.1 และโครงสร้างที่ซับซ้อนมากขึ้น (MS) ดังตารางที่ 3.2 ปรากฏขนาดของมุมในมิติที่ 1 (ข้อ 1-15) และมิติที่ 2 (ข้อ 16-30) ดังรายละเอียดในตารางที่ 3.5

ตารางที่ 3.5 ขนาดของมุมในมิติที่ 1 และมิติที่ 2 ซึ่งแยกตามโครงสร้างที่ไม่ซับซ้อน (APSS) และโครงสร้างที่มีความซับซ้อน (MS) เพื่อใช้สำหรับการสร้างค่าพารามิเตอร์อำนาจจำแนก (a_1 และ a_2) และค่าความยากของข้อสอบ (d) สำหรับการแปลงคะแนนในทุกระดับชั้น

ลำดับข้อสอบที่จัดอยู่ในมิติที่ 1	APSS		MS	
	ขนาดของมุมจากแกน θ_1	ขนาดของมุมจากแกน θ_2	ขนาดของมุมจากแกน θ_1	ขนาดของมุมจากแกน θ_2
1	1.00	89.00	0.50	89.50
2	1.07	88.93	2.50	87.50
3	2.14	87.86	5.00	85.00
4	3.21	86.79	7.50	82.50
5	4.29	85.71	10.00	80.00
6	5.36	84.64	12.50	77.50
7	6.43	83.57	15.00	75.00
8	7.50	82.50	25.00	65.00
9	8.57	81.43	26.88	63.13
10	9.64	80.36	28.75	61.25
11	10.71	79.29	30.63	59.38
12	11.79	78.21	32.50	57.50
13	12.86	77.14	34.38	55.63
14	13.93	76.07	36.25	53.75
15	15.00	75.00	40.00	50.00
MEAN	7.57	82.43	20.49	69.51
SD	4.69	4.69	13.45	13.45
ลำดับข้อสอบที่จัดอยู่ในมิติที่ 2	**APSS**		**MS**	
	ขนาดของมุมจากแกน θ_1	ขนาดของมุมจากแกน θ_2	ขนาดของมุมจากแกน θ_1	ขนาดของมุมจากแกน θ_2
16	75.00	15.00	50.00	40.00
17	76.07	13.93	51.88	38.13
18	77.14	12.86	53.75	36.25
19	78.21	11.79	55.63	34.38
20	79.29	10.71	57.50	32.50
21	80.36	9.64	59.38	30.63
22	81.43	8.57	61.25	28.75
23	82.50	7.50	65.00	25.00
24	83.57	6.43	75.00	15.00
25	84.64	5.36	77.50	12.50
26	85.71	4.29	80.00	10.00
27	86.79	3.21	82.50	7.50
28	87.86	2.14	85.00	5.00
29	88.93	1.07	87.50	2.50
30	89.00	1.00	89.50	0.50
MEAN	82.43	7.57	68.76	21.24
SD	38.31	38.39	27.98	28.06

จากตารางที่ 3.5 เมื่อพิจารณาตามโครงสร้างมิติความสามารถ พบว่า โครงสร้างที่ไม่ซับซ้อน (APSS) ในมิติที่ 1 มีค่าเฉลี่ยของขนาดของมุม 15 ข้อแรกและ 15 ข้อหลัง ประมาณ 7.57° และ 82.43° ตามลำดับ ส่วนมิติที่ 2 ผลที่ตรงข้ามกัน เมื่อพิจารณาโครงสร้างที่ซับซ้อนมากขึ้น (MS) พบว่ามีค่าเฉลี่ยของขนาดของมุมในมิติที่ 1 เป็น 4 ช่วง พบว่ามีขนาดของมุมโดยเฉลี่ย เท่ากับ 7.57° 35.78°, 55.63° และ 80.25° ตามลำดับ

นอกจากนี้เมื่อทราบค่าความยากรวม (MDIFF) ที่แสดงในตารางที่ 3.3 และ 3.4 สามารถคำนวณค่า d_i ได้โดยตรง ดังแสดงในสมการ

$$MDIFF_i = \frac{-d_i}{MDISC_i} \quad \text{.....(3.9)}$$

เมื่อ *MDIFF* แทน ค่าอำนาจจำแนกรวมของข้อสอบข้อที่ i ในมิติที่ m

เมื่อ d_i มีความสัมพันธ์กับค่าความยากของแบบสอบ ซึ่ง $MDIFF_i$ เป็นระยะทางจากจุดตำแหน่งมิติของ Θ (Θ-space) ซึ่งเป็นจุดที่มีความชันสูงสุด

สำหรับชุดของค่าพารามิเตอร์ของข้อสอบร่วม (a_{ik} และ d_i) จำนวน 20 ข้อ เพื่อใช้ในการสร้างรูปแบบการตอบสนองแบบสอบในแต่ละครั้งที่ทำการศึกษา ผู้วิจัยจะทำการศึกษาโดยมีการทำซ้ำจำนวน 40 ครั้ง (40 Replication) โดยแบ่งเป็นการศึกษาของโครงสร้างที่ไม่ซับซ้อน จำนวน 20 ครั้ง และโครงสร้างที่ซับซ้อนมากขึ้น (MS) จำนวน 20 ครั้ง

สำหรับงานวิจัยในครั้งนี้ผู้วิจัยทำการคำนวณค่าพารามิเตอร์ความยากในมิติที่ 1 และมิติที่ 2 โดยในแต่ละโครงสร้างจะมีขนาดของมุมแตกต่างกันไปดังที่ได้กล่าวมาแล้วข้างต้น โดยอาศัยหลักการของการประมาณค่าพารามิเตอร์การแปลงคะแนน (Reckase, 2007) ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ (a_1 และ a_2) และค่าความยาก (d) ของข้อสอบร่วม ซึ่งคำนวณได้จากสมการที่ 3.9 ซึ่งได้จากผลคูณของ MDISC และ MDIFF จากแบ่งช่วงการกระจายของค่าพารามิเตอร์ MDISC และ MDIFF 5 ระดับ ที่ศึกษาโดย Yon (2007) ดังตารางที่ 3.3 และ 3.4 ซึ่งมีข้อสอบร่วมจำนวน 30 ข้อ ของนักเรียนในระดับที่ 1 ทำแบบสอบในระดับที่ 2 และนักเรียนในระดับที่ 2 ทำแบบสอบในระดับที่ 2 ซึ่งนักเรียนในระดับที่ 1 ทำแบบสอบในระดับที่ 2 จะมีค่า MDIFF มากกว่านักเรียนในระดับที่ 2 ทำแบบสอบในระดับที่ 2 ที่เหมาะสำหรับโครงสร้างที่ไม่ซับซ้อน (APSS) และโครงสร้างที่ซับซ้อนมากขึ้น (MS) เพื่อใช้สำหรับการแปลงคะแนนจากระดับที่ 1 (L1) ไปยังระดับที่ 2 (L2) สำหรับค่าพารามิเตอร์เพื่อใช้สำหรับการแปลงคะแนนจากระดับที่ 2 (L3) ไปยังระดับที่ 3 (L3) อาศัยหลักการเดียวกัน ซึ่งผลการสร้างค่าพารามิเตอร์ที่แท้จริงเมื่อพิจารณาในรายละระดับของการแปลงคะแนน ปรากฏผลดังตารางที่ 3.6 และ 3.7 ตามลำดับ

ตารางที่ 3.6 ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ (a_1 และ a_2) และค่าความยาก (d) ของข้อสอบร่วมจำนวน 30 ข้อ ของนักเรียนในระดับที่ 1 ทำแบบสอบในระดับที่ 2 และนักเรียนในระดับที่ 2 ทำแบบสอบในระดับที่ 2 ที่เหมาะสำหรับโครงสร้างที่ไม่ซับซ้อน (APSS) และโครงสร้างที่ซับซ้อนมากขึ้น (MS) เพื่อใช้สำหรับการแปลงคะแนนจากระดับที่ 1 (L1) ไปยังระดับที่ 2 (L2)

item	APSS		MS		L1 Take L2 Test		L2 Take L2 Test		MDISC
	a1	a2	a1	a2	d_{12}	$MDIFF_{12}$	d_2	$MDIFF_2$	
1	1.30	0.02	1.30	0.01	-3.90	3.00	-2.86	2.20	1.30
2	1.20	0.02	1.20	0.05	-0.60	0.50	0.00	0.00	1.20
3	1.00	0.04	1.00	0.09	-1.50	1.50	-1.00	1.00	1.00
4	0.80	0.04	0.79	0.10	-2.00	2.50	-1.60	2.00	0.80
5	0.60	0.04	0.59	0.10	0.00	0.00	0.30	-0.50	0.60
6	1.29	0.12	1.27	0.28	-3.90	3.00	-2.86	2.20	1.30
7	1.19	0.13	1.16	0.31	-0.60	0.50	0.00	0.00	1.20
8	0.99	0.13	0.91	0.42	-1.50	1.50	-1.00	1.00	1.00
9	0.79	0.12	0.71	0.36	-2.00	2.50	-1.60	2.00	0.80
10	0.59	0.10	0.53	0.29	0.00	0.00	0.30	-0.50	0.60
11	1.28	0.24	1.12	0.66	-3.90	3.00	-2.86	2.20	1.30
12	1.17	0.25	1.01	0.64	-0.60	0.50	0.00	0.00	1.20
13	0.97	0.22	0.83	0.56	-1.50	1.50	-1.00	1.00	1.00
14	0.78	0.19	0.65	0.47	-2.00	2.50	-1.60	2.00	0.80
15	0.58	0.16	0.46	0.39	0.00	0.00	0.30	-0.50	0.60
16	0.34	1.26	0.84	1.00	-3.90	3.00	-2.86	2.20	1.30
17	0.29	1.16	0.74	0.94	-0.60	0.50	0.00	0.00	1.20
18	0.22	0.97	0.59	0.81	-1.50	1.50	-1.00	1.00	1.00
19	0.16	0.78	0.45	0.66	-2.00	2.50	-1.60	2.00	0.80
20	0.11	0.59	0.32	0.51	0.00	0.00	0.30	-0.50	0.60
21	0.22	1.28	0.66	1.12	-3.90	3.00	-2.86	2.20	1.30
22	0.18	1.19	0.58	1.05	-0.60	0.50	0.00	0.00	1.20
23	0.13	0.99	0.42	0.91	-1.50	1.50	-1.00	1.00	1.00
24	0.09	0.79	0.21	0.77	-2.00	2.50	-1.60	2.00	0.80
25	0.06	0.60	0.13	0.59	0.00	0.00	0.30	-0.50	0.60
26	0.10	1.30	0.23	1.28	-3.90	3.00	-2.86	2.20	1.30
27	0.07	1.20	0.16	1.19	-0.60	0.50	0.00	0.00	1.20
28	0.04	1.00	0.09	1.00	-1.50	1.50	-1.00	1.00	1.00
29	0.01	0.80	0.03	0.80	-2.00	2.50	-1.60	2.00	0.80
30	0.01	0.60	0.01	0.60	0.00	0.00	0.30	-0.50	0.60
MEAN	0.55	0.54	0.63	0.60	-1.60	1.50	-1.03	0.94	0.98
SD	0.47	0.47	0.38	0.36	1.37	1.16	1.16	1.08	0.26

ตารางที่ 3.7 ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ (a_1 และ a_2) และค่าความยาก (d) ของข้อสอบร่วมจำนวน 30 ข้อ ของนักเรียนในระดับที่ 2 ทำแบบสอบในระดับที่ 3 และนักเรียนในระดับที่ 3 ทำแบบสอบในระดับที่ 3 ที่เหมาะสำหรับโครงสร้างที่ไม่ซับซ้อน (APSS) และโครงสร้างที่ซับซ้อนมากขึ้น (MS) เพื่อใช้สำหรับการแปลงคะแนนจากระดับที่ 2 (L2) ไปยังระดับที่ 3 (L3)

item	APSS		MS		L2 Take L3 Test		L3 Take L3 Test		MDISC
	a1	a2	a1	a2	d_{23}	$MDIFF_{23}$	d_3	$MDIFF_3$	
1	1.30	0.02	1.30	0.01	-3.90	3.00	-2.86	2.20	1.30
2	1.00	0.02	1.00	0.04	-1.50	1.50	-1.00	1.00	1.00
3	0.80	0.03	0.80	0.07	-2.00	2.50	-1.60	2.00	0.80
4	1.30	0.07	1.29	0.17	-3.90	3.00	-2.86	2.20	1.30
5	0.80	0.06	0.79	0.14	-2.00	2.50	-1.60	2.00	0.80
6	1.29	0.12	1.27	0.28	-3.90	3.00	-2.86	2.20	1.30
7	0.99	0.11	0.97	0.26	-1.50	1.50	-1.00	1.00	1.00
8	0.79	0.10	0.73	0.34	-2.00	2.50	-1.60	2.00	0.80
9	1.29	0.19	1.16	0.59	-3.90	3.00	-2.86	2.20	1.30
10	0.59	0.10	0.53	0.29	0.00	0.00	0.30	-0.50	0.60
11	1.28	0.24	1.12	0.66	-3.90	3.00	-2.86	2.20	1.30
12	0.98	0.20	0.84	0.54	-1.50	1.50	-1.00	1.00	1.00
13	0.78	0.18	0.66	0.45	-2.00	2.50	-1.60	2.00	0.80
14	1.26	0.31	1.05	0.77	-3.90	3.00	-2.86	2.20	1.30
15	1.16	0.31	0.92	0.77	-0.60	0.50	0.00	0.00	1.20
16	0.34	1.26	0.84	1.00	-3.90	3.00	-2.86	2.20	1.30
17	0.24	0.97	0.62	0.79	-1.50	1.50	-1.00	1.00	1.00
18	0.18	0.78	0.47	0.65	-2.00	2.50	-1.60	2.00	0.80
19	0.27	1.27	0.73	1.07	-3.90	3.00	-2.86	2.20	1.30
20	0.22	1.18	0.64	1.01	-0.60	0.50	0.00	0.00	1.20
21	0.22	1.28	0.66	1.12	-3.90	3.00	-2.86	2.20	1.30
22	0.15	0.99	0.48	0.88	-1.50	1.50	-1.00	1.00	1.00
23	0.10	0.79	0.34	0.73	-2.00	2.50	-1.60	2.00	0.80
24	0.15	1.29	0.34	1.26	-3.90	3.00	-2.86	2.20	1.30
25	0.11	1.19	0.26	1.17	-0.60	0.50	0.00	0.00	1.20
26	0.10	1.30	0.23	1.28	-3.90	3.00	-2.86	2.20	1.30
27	0.04	0.80	0.10	0.79	-2.00	2.50	-1.60	2.00	0.80
28	0.03	0.80	0.07	0.80	-2.00	2.50	-1.60	2.00	0.80
29	0.02	1.30	0.06	1.30	-3.90	3.00	-2.86	2.20	1.30
30	0.01	0.60	0.01	0.60	0.00	0.00	0.30	-0.50	0.60
MEAN	0.59	0.60	0.68	0.66	-2.40	2.17	-1.72	1.55	1.06
SD	0.49	0.50	0.38	0.39	1.36	1.00	1.10	0.91	0.24

สำหรับการประมาณค่าพารามิเตอร์ของข้อสอบในโมเดล MIRT ความน่าจะเป็นในการตอบ (P_{ij}) จากผู้สอบแต่ละคน สามารถคำนวณได้จากโมเดลที่การประมาณค่าความสามารถ (Θ) ในมิติต่างกันสามารถชดเชยกันได้ (Compensatory Model) ซึ่งเป็นโมเดลโลจิสติกชนิด 3 พารามิเตอร์ (McKinly & Reckase, 1983; Reckase, 1985, 1995) ดังแสดงในสมการที่ 3.10

$$P(U_{ij}=1|a_i,d_i,\theta_j,c_i)=c_i+(1-c_i)\frac{\exp(a_i'\theta_j+d_i}{1+\exp(a_i'\theta_j+d_i} \quad \text{.....(3.10)}$$

เมื่อ U_{ij} แทน คะแนน (0/1) สำหรับข้อสอบข้อที่ i ของคนที่ j

a_i แทน เวคเตอร์ของค่าพารามิเตอร์ที่มีความสัมพันธ์กับอำนาจจำแนกของข้อสอบข้อที่ i ในแต่ละมิติที่มีจำนวน m มิติ

D แทน ค่าพารามิเตอร์ที่มีลักษณะเป็นค่าสเกลลาที่มีความสัมพันธ์กับค่าความยากของข้อสอบข้อที่ i

θ_j แทน เวคเตอร์ของค่าพารามิเตอร์ที่อธิบายความสามารถของผู้สอบคนที่ j ในมิติที่ m (m-Dimension Space)

c_i แทน ค่าพารามิเตอร์การเดาของข้อสอบ นั่นคือเป็นค่าความน่าจะเป็นที่บุคคลที่มีความสามารถต่ำจะตอบข้อสอบได้ถูกต้อง

เมื่อพิจารณาสมการดังกล่าวข้างต้น $P(U_{ij}=1|a_i,d_i,\theta_j,c_i)$ คือความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง (ได้ 1 คะแนน) สำหรับข้อสอบข้อที่ i สำหรับบุคคลที่ j ในมิติที่ m ซึ่งความน่าจะเป็นในการตอบข้อสอบมีค่าต่ำแสดงว่าข้อสอบนั้นมีความยาก และ/ หรือ ความสามารถของผู้สอบอยู่ในระดับต่ำสำหรับข้อสอบข้อนั้น ซึ่งผลที่ได้จากการคำนวณ P_{ij} จะสามารถเปรียบเทียบกับการสุ่มแบบสม่ำเสมอ (Uniform Random) ซึ่งเรียกว่า P^* ที่มีการให้คะแนนแบบ 2 ค่า (Binary Item Score) ผลการเปรียบเทียบกำหนดให้ $x_{ij}=1$ เมื่อ $P_{ij} > P^*$ เพื่อบ่งชี้ว่าตอบข้อคำถามได้ถูกต้อง ในทางกลับกัน กำหนดให้คะแนนของ $x_{ij} = 0$ เมื่อ $P^* \geq P$

ในการสร้างลักษณะการแจกแจงความสามารถ (Generate Ability Distribution) ในครั้งนี้จะจำลองข้อมูลให้มีลักษณะการแจกแจงปกติแบบแบ่งสอง (Bivariate Normal Distribution) เพื่อให้ค่าเฉลี่ยและค่าความแปรปรวน-ความแปรปรวนร่วมอย่างถูกต้อง ซึ่งในที่นี้ได้ใช้โปรแกรม MATLAB ในการจำลองข้อมูล

กล่าวโดยสรุปในขั้นตอนนี้เป็นการสร้างรูปแบบการตอบสนองข้อสอบที่มีลักษณะของการตรวจให้คะแนนแบบ 2 ค่า นั่นคือ ตอบถูกได้ 1 คะแนน และตอบผิดได้ 0 คะแนน ตามโมเดล MIRT ดังสมการที่ 3.3 โดยใช้โปรแกรม MATLAB ในการสร้างข้อมูลขึ้นมา โดยกำหนดให้ค่าพารามิเตอร์ของข้อสอบทั้ง 3 ระดับชั้น มีลักษณะการกระจายความสามารถ ที่มีโครงสร้างแตกต่างกัน 2 ลักษณะ คือโครงสร้างที่ไม่ซับซ้อน (APSS) และโครงสร้างที่ซับซ้อนมากขึ้น (MS) รวมถึงค่าความสามารถเฉลี่ยและค่าความแปรปรวน-ความแปรปรวนร่วม

การสร้างสเกลการปรับเทียบแนวตั้งที่แท้จริง (Generation of a True Vertical Scale)

ในการสร้างสเกลการปรับเทียบแนวตั้ง คุณสมบัติและชุดการกระจายความสามารถของผู้สอบจะถูกสร้างขึ้นให้อยู่ในแกน (Reference) ที่สามารถเปรียบเทียบกันได้สำหรับในสถานการณ์จริงแล้ว ความเป็นตัวแทนของการกระจายความสามารถของผู้สอบในการศึกษาครั้งนี้จะศึกษากับกลุ่มผู้สอบที่มีความสามารถแตกต่างกัน 3 ระดับชั้น ในที่นี้ยกตัวอย่างเป็นระดับที่ 1, 2 และ 3 ตามลำดับ ซึ่งความสามารถที่แท้จริงจะถูกสร้างขึ้นจากกลุ่มตัวอย่างของนักเรียนในแต่ละระดับชั้นจากข้อมูลที่มีการแจกแจงปกติแบบแบ่งสอง (Bivariate Normal Distribution) หรือ N(μ_g, σ_g) เมื่อ *g* คือ ระดับชั้นที่ *g เช่น* ระดับชั้นที่ 1,2 หรือ 3 เป็นต้น โดยมีระดับความสัมพันธ์ของมิติความสามารถในแต่ละระดับชั้นที่ทำการศึกษาแบ่งออกเป็น 3 ระดับ คือไม่มีระดับความสัมพันธ์ ความสัมพันธ์ระดับกลาง และความสัมพันธ์ระดับสูง ดังแสดงในตารางที่ 3.8

ตารางที่ 3.8 การกระจายค่าความสามารถสำหรับผู้สอบแต่ละกลุ่ม

ระดับชั้น (Level)	เวคเตอร์ของค่าเฉลี่ย (μ)	ค่าเมตริกซ์ความแปรปรวน -ความแปรปรวนร่วม (Σ)	ระดับความสัมพันธ์ของมิติความสามารถ (ρ)
ระดับที่ 1 (ระดับที่ 1 ทำแบบสอบระดับที่ 2)	$\begin{bmatrix}0\\0\end{bmatrix}$	$\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$	ρ =0.0
			ρ =0.5
			ρ =0.8
ระดับที่ 2 (ระดับที่ 2 ทำแบบสอบระดับที่ 2)	$\begin{bmatrix}0.5\\0.5\end{bmatrix}$	$\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$	ρ =0
			ρ =0.5
			ρ =0.8
ระดับที่ 2 (ระดับที่ 2 ทำแบบสอบระดับที่ 3)	$\begin{bmatrix}0.5\\0.5\end{bmatrix}$	$\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$	ρ =0.0
			ρ =0.5
			ρ =0.8
ระดับที่ 3 (ระดับที่ 3 ทำแบบสอบระดับที่ 3)	$\begin{bmatrix}1\\1\end{bmatrix}$	$\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$	ρ =0.0
			ρ =0.5
			ρ =0.8

โดยทั่วไปแล้วค่าเฉลี่ย (μ_g) จะมีค่าเพิ่มขึ้นเมื่อระดับชั้นสูงขึ้น และค่าความแปรปรวนของแต่ละระดับชั้นจะเป็นค่าคงที่ร่วมกัน ซึ่งการเพิ่มขึ้นของค่าเฉลี่ยในทั้ง 2 มิติจะครอบคลุมทุกระดับชั้น ภายใต้สมมติฐานที่ทำการศึกษาเกี่ยวกับพัฒนาการ (Growth Study) ที่ทำการศึกษาโดย Tate (2003) ผลการวิจัย พบว่า ความสามารถโดยเฉลี่ยของประชากรจะมีแนวโน้มสูงขึ้นเมื่ออยู่ในระดับชั้นที่สูงขึ้นตลอดช่วงที่ศึกษา ซึ่งเป็นการศึกษาในระยะยาว (Longitudinal Study)

ของนักเรียนในอเมริกา (LSAY) นอกจากนี้ Ma & Wilkins (2002) ได้ศึกษาโมเดลเชิงเส้นตรงแบบลดหลั่น (Hierarchical Linear Model) ใน 3 มิติ คือ ชีววิทยา (Biology) วิทยาศาสตร์กายภาพ (Physical Science) และวิทยาศาสตร์สิ่งแวดล้อม (Environmental Science) ในโรงเรียนประถมและมัธยมศึกษา ผลที่ได้แสดงให้เห็นว่า มีความแตกต่างของพัฒนาการ ในแต่ละระดับชั้นอย่างมีนัยสำคัญทางสถิติ ซึ่งมีพัฒนาการเพิ่มสูงขึ้นอย่างรวดเร็วสำหรับนักเรียนระดับประถมศึกษา แต่มีพัฒนาการค่อนข้างช้าในระดับมัธยมศึกษา โดยเฉพาะในระดับชั้นที่ใกล้จะจบการศึกษา

การปรับเทียบสเกลแนวตั้ง (Vertical Scaling) จะถูกจำลองขึ้นตามการออกแบบการใช้ข้อสอบร่วมจากกลุ่มที่ไม่เท่าเทียมกัน (Common-Item Non-Equivalent Group Design) โดยการออกแบบในครั้งนี้ นักเรียนในแต่ละระดับชั้นจะทำแบบสอบในระดับชั้นของตนเองและแบบสอบในระดับชั้นสูงกว่าที่อยู่ติดกัน กำหนดให้นักเรียนในระดับที่ 1 (n=2,000 คน) ทำแบบสอบระดับที่ 2 ส่วนนักเรียนในระดับที่ 2 (n=2,000 คน) ทำแบบสอบในระดับตนเองและทำแบบสอบระดับที่ 3 ส่วนนักเรียนระดับที่ 3 ทำแบบสอบในระดับชั้นของตนเอง การจำลองข้อมูลดังกล่าวสามารถแสดงได้ดังตารางที่ 3.9

ตารางที่ 3.9 การออกแบบการเชื่อมโยงคะแนนสำหรับการจำลองข้อมูล: การใช้ข้อสอบจากกลุ่มผู้สอบที่ไม่เท่าเทียมกัน (Commom-Item Non-Equvalent Group Design)

ระดับของแบบสอบ	จำนวนข้อสอบร่วม	จำนวนนักเรียนในแต่ละระดับชั้น		
		ระดับที่ 1	ระดับที่ 2	ระดับที่ 3
ระดับที่ 1	30	2,000		
ระดับที่ 2	30		2,000	
ระดับที่ 3	30			2,000

2.2 การประมาณค่าพารามิเตอร์ตามโมเดล MIRT (MIRT Calibration)

ภายหลังทำการสร้างรูปแบบการตอบข้อสอบ จากนั้นนำค่าพารามิเตอร์ของข้อสอบแต่ละข้อที่ได้จากขั้นตอนที่ 1 ตามสมการที่ 3.10 จะ Calibrated ข้อมูลตามโมเดล MIRT โดยใช้โปรแกรม TESTFACT ที่พัฒนาโดย Wilson, Wood & Gibbons (1984) โดยใช้วิธีการที่ไม่ใช่เชิงเส้นตรง มีลักษณะของการวิเคราะห์องค์ประกอบแบบสำรวจชนิดเต็มรูปแบบ (Non-Linear, Exploratory Full Information Factor Analysis) หลังจากนั้นจะใช้โปรแกรม BILOG-MG (Zimowski, Muraki, Mislevy & Bock, 2003) ในการประมาณค่าพารามิเตอร์โอกาสการเดาข้อสอบ (Lower Asymptote Parameter; c_i)

TESTFACT เป็นโปรแกรมคอมพิวเตอร์ที่ออกแบบขึ้นเพื่อใช้กับข้อมูลที่มีการตรวจให้คะแนน 2 ค่า คือ 1 และ 0 โดยมีข้อตกลงเบื้องต้นว่ากลุ่มตัวอย่างได้มาจากการสุ่มของประชากร และ

ระดับความสามารถของคุณลักษณะแฝง (Latent Trait) ของผู้สอบมาจากการแจกแจงปกติแบบพหุ (MVN (0,1)) (Mislevy, 1986) นั่นคือมีค่าเฉลี่ยเป็น 0 และมีส่วนเบี่ยงเบนมาตรฐานเป็น 1 (Knol & Berger, 1988) ซึ่งโปรแกรมที่ใช้ในการประมาณค่าโมเดล MIRT ส่วนใหญ่รวมถึงโปรแกรม TESTFACT ได้มีการแสดงเพื่อระบุปัญหา (หรือการสร้างวิธีการที่เป็นแบบแผนเดียวกัน) คือ เหมาะสมกับลักษณะของข้อมูลที่มีการแจกแจงปกติแบบพหุมิติ โดย Reckase (1997) ระบุว่า *"ถ้าความสัมพันธ์ระหว่างแต่ละมิติมีค่าคงที่ซึ่งมีค่าเท่ากับ 0 ดังนั้นความสัมพันธ์ที่สามารถสังเกตได้ระหว่างคะแนนของข้อสอบแต่ละข้อ จะพิจารณาได้จากค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ (a-parameter) นั่นเอง ทำนองเดียวกันถ้าค่าความแปรปรวนของความสามารถแบบพหุมิติมีค่าคงที่ ซึ่งมีค่าเท่ากับ 1 การประมาณค่าพารามิเตอร์อำนาจจำแนกสามารถได้จากการวิเคราะห์โดยใช้โปรแกรมคอมพิวเตอร์ที่วิเคราะห์โมเดล MIRT โดยทั่วไป ที่จะมีการใช้ค่าความแปรปรวนของจุดกำเนิดที่มีหลายมิติ (Original Heterogeneous Variances) ดังนั้นการกำหนดค่า Θ ซึ่งมีการแจกแจงปกติแบบพหุ ตามลักษณะของค่าความแปรปรวนและค่าความแปรปรวนร่วมของจุดกำเนิดที่มีหลายมิติ จะยึดตามค่าความแปรปรวนความแปรปรวนร่วมของค่าพารามิเตอร์อำนาจจำแนก (a) ในพหุมิติ"*

นอกจากนี้ผลลัพธ์ที่ได้จากโปรแกรม TESTFACT มีการนำเสนอค่าเฉลี่ยของส่วนที่เหลือ (Mean Residual Values) ซึ่งมีความสอดคล้องกับจำนวนของการแปรผันชุดของข้อมูลที่ไม่สามารถวัดได้จากมิติที่ทำการศึกษาในโมเดล ตามที่ขนาดค่าเฉลี่ยของส่วนที่เหลือ สามารถนำมาใช้ในการตรวจสอบความเพียงพอของโมเดลที่ใช้สำหรับข้อมูลจากการจำลองขึ้นมา โดยถ้าค่าเฉลี่ยของส่วนที่เหลือมีค่ามาก ซึ่งโดยทั่วไปมีค่ามากกว่า 0.05 แสดงว่าข้อมูลที่ได้ไม่สามารถอธิบายค่าความแปรผันของชุดของข้อมูลได้อย่างเพียงพอ ดังนั้นในการศึกษาครั้งนี้คาดหวังว่าโมเดลที่ศึกษาทั้ง 2 มิติมีความเพียงพอและสอดคล้องกับโมเดล MIRT ที่ศึกษาสำหรับข้อมูลที่มีการจำลองข้อมูลขึ้นมา อย่างไรก็ตามค่าเฉลี่ยของส่วนที่เหลือ (Mean Residual) ที่ได้จากโปรแกรม TESFACT ควรมีค่าน้อยกว่า 0.05 จึงถือว่าข้อมูลสอดคล้องกับโมเดล (Yon, 2006, 2007) ซึ่งเป็นค่าที่ยอมรับได้ และมีความเพียงพอในการวิเคราะห์ตามสมมติฐานที่ตั้งไว้ เพื่ออธิบายรูปแบบการตอบในการจำลองข้อมูล และเพื่อให้สอดคล้องกับโครงสร้าง 2 มิติที่คาดหวังไว้

2.3 การเชื่อมโยงคะแนนหรือการปรับเทียบคะแนน (Equating Procedure)

การศึกษาครั้งนี้ เป็นการประมาณค่าลักษณะของข้อสอบที่ใช้ข้อสอบร่วม 30 ข้อ ซึ่งจะถูกแปลงไปยังแบบสอบฐาน ด้วยวิธีการปรับเทียบทั้ง 2 วิธี โดยกำหนดให้แบบสอบในระดับที่ 3 เป็นแบบสอบฐานส่วนแบบสอบในระดับชั้นที่ต่ำกว่าเพื่อนำมาเปรียบเทียบกับผลที่ได้กำหนดให้เป็นแบบสอบปรับเทียบคะแนน การเลือกแบบสอบในระดับชั้นสูงสุดเป็นแบบสอบฐาน เนื่องจากผลสัมฤทธิ์ของนักเรียนมีหลายมิติที่ศึกษาในรายวิชาเดียวกันมีความแตกต่างของค่า

ความสามารถในระดับชั้นที่สูงกว่ามากกว่าระดับชั้นที่ต่ำกว่า (Reckase & Martineau, 2004) โดยการศึกษาในครั้งนี้จะประมาณค่าพารามิเตอร์ของผลที่ได้จากนักเรียนระดับที่ 2 ที่ทำแบบสอบระดับที่ 3 จะถูกปรับเทียบค่าพารามิเตอร์ไปยังแบบสอบของนักเรียนระดับที่ 3 ที่ทำแบบสอบในระดับที่ 3 แบบสอบระดับที่ 2 ไปยังระดับที่ 3 จะเป็นการปรับเทียบค่าพารามิเตอร์ การแปลงคะแนนที่มีลักษณะคู่ขนานจะประยุกต์ใช้ในการประมาณค่าพารามิเตอร์ หลังจากนั้นทำการปรับเทียบแนวตั้งใน 2 ระดับชั้น ซึ่งสามารถเปรียบเทียบกันได้โดยตรง สุดท้ายการประมาณค่าพารามิเตอร์ที่สร้างจากนักเรียนในระดับที่ 1 ไปยังระดับที่ 2 ดังแสดงในแผนภาพที่ 3.4

การประมาณค่าพารามิเตอร์ของแบบสอบในระดับชั้นที่อยู่ต่ำกว่าจะถูกแปลงคะแนนไปยังเมตริกซ์ของแบบสอบในระดับชั้นที่อยู่สูงกว่าตามวิธีการปรับเทียบในแต่ละวิธีที่ศึกษา นั่นคือ วิธี M และวิธี NOP โดยในครั้งนี้จะทำการศึกษาเพียง 1 ครั้ง (Replication) ในแต่ละวิธี โดยใช้โปรแกรม MATLAB ในการเชื่อมโยงคะแนน

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตอนที่ 3 การวิเคราะห์ข้อมูลที่ได้จากการจำลองข้อมูลและเกณฑ์ที่ใช้ในการเปรียบเทียบคุณภาพของการเชื่อมโยงคะแนน (Analysis of Simulation Study and Criteria for Comparison of Quality)

ในโครงสร้างการวิเคราะห์ตามทฤษฎี IRT เกณฑ์หนึ่งที่สำคัญเพื่อประเมินการเชื่อมโยงคะแนนโดยใช้ข้อสอบร่วม คือพิจารณาขนาดความคลาดเคลื่อนที่เกิดจากวิธีการเชื่อมโยงคะแนน โดยคำนวณค่าความแตกต่างระหว่างการประมาณจากแบบสอบฐาน และการประมาณค่าการแปลงคะแนน ที่ได้จากแบบสอบปรับเทียบคะแนน รวมถึงการพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างการประมาณค่าพารามิเตอร์จากแบบสอบฐานและค่าการแปลงคะแนน ที่ได้จากแบบสอบปรับเทียบคะแนน ซึ่งมีรายละเอียดของเกณฑ์ในการพิจารณาดังนี้

3.1 ค่าความลำเอียง (BIAS) นั่นคือ การประมาณค่าที่ได้จากการแปลงคะแนนของแบบสอบปรับเทียบคะแนนมีระยะทางที่ต่างจากค่าพารามิเตอร์ที่แท้จริงของข้อสอบโดยเฉลี่ยมากน้อยเพียงใด ซึ่งสามารถแสดงได้ดังสมการที่ 3.11

3.2 ค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (Root-Mean Square Error; RMSE) เป็นขนาดของความแตกต่างระหว่างการประมาณค่าจากแบบสอบปรับเทียบคะแนนและแบบสอบฐาน นั่นคือ มีความแตกต่างทั้งหมดของการแปลงคะแนนข้ามกลุ่มข้อสอบร่วม (Common Items) มากน้อยเพียงใด ซึ่งสามารถแสดงได้ดังสมการที่ 3.12

$$BIAS = \sum_{i=1}^{I} \frac{\left(a_{1i}^{*} - a_{1i}\right)}{I} \quad \text{.....(3.11)}$$

$$RMSE = \left[\frac{\sum_{i=1}^{I} \left(a_{1i}^{*} - a_{1i}\right)^2}{I-1} \right]^{1/2} \quad \text{.....(3.12)}$$

เมื่อ a_{1i} แทน ค่าพารามิเตอร์อำนาจจำแนก (Discrimination Parameter) ของมิติที่ 1 สำหรับข้อสอบข้อที่ i

a^{*}_{1i} แทน ค่าอำนาจจำแนกที่ได้จากการแปลงคะแนน (Transformed Discrimination)

I แทน จำนวนของข้อสอบร่วม

3.3 ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) ซึ่งเป็นค่าที่แสดงความสัมพันธ์ระหว่างการประมาณค่าที่ได้จากการแปลงคะแนนจากแบบสอบปรับเทียบคะแนนและการประมาณค่าเริ่มแรกของแบบสอบฐาน

3.4 ปฏิสัมพันธ์ของวิธีการเชื่อมโยงคะแนน โครงสร้างมิติความสามารถ และระดับความสัมพันธ์ระหว่างมิติความสามารถ พิจารณาจากค่าที่ได้จากการวิเคราะห์ความแปรปรวนแบบพหุจำแนก 3 ทาง (Three-way MANOVA)) เพื่อศึกษาปฏิสัมพันธ์ระหว่างวิธีการเชื่อมโยงคะแนน โครงสร้างมิติความสามารถและระดับความสัมพันธ์ระหว่างมิติความสามารถ ซึ่งส่งผลต่อคุณภาพของการเชื่อมโยงคะแนนที่พิจารณาจากค่าดัชนี RMSE BIAS และ CORR ของการแปลงค่าพารามิเตอร์อำนาจจำแนกในมิติที่ 1 (a_1) และมิติที่ 2 (a_2) โดยที่แต่ละตัวแปรต้นแบ่งเป็นหลายระดับคือ 1) วิธีการเชื่อมโยงคะแนนแบ่งเป็น 2 ระดับ คือ วิธี M และ วิธี NOP 2) โครงสร้างมิติความสามารถแบ่งเป็น 2 ระดับ คือ โครงสร้างที่ไม่ซับซ้อน (APSS) และโครงสร้างที่ซับซ้อนมากขึ้น (MS) และ 3) ความสัมพันธ์ระหว่างมิติความสามารถแบ่งเป็น 3 ระดับ คือ ไม่มีระดับความสัมพันธ์ (r= 0.0) ความสัมพันธ์ระดับกลาง (r=0.5) และความสัมพันธ์ระดับสูง (r=0.8)

ผู้วิจัยวิเคราะห์ข้อมูลในส่วนนี้เพื่อตรวจสอบความแตกต่างของเซ็นทรอยด์ ซึ่งเป็นความแตกต่างระหว่างค่าเฉลี่ยของของตัวแปรดัชนีบ่งชี้คุณภาพ RMSE BIAS และ CORR ของค่าพารามิเตอร์อำนาจจำแนกในมิติที่ 1 และ มิติที่ 2 ซึ่งเป็นตัวแปรตามทั้งหมด 6 ตัวแปร คือ 1) ค่าดัชนี Ln RMSE ในมิติที่ 1 (Ln RMSE_a1) 2) ค่าดัชนี Ln RMSE ในมิติที่ 2 (Ln RMSE_a2) 3) ค่าดัชนี BIAS ในมิติที่ 1 (BIAS_a1) 4) ค่าดัชนี BIAS ในมิติที่ 2 (BIAS_a2) 5) ค่า CORR ในมิติที่ 1 (CORR_a_1) และ 6) ค่า CORR ในมิติที่ 2 (CORR_a_2)

ตอนที่ 4 การตรวจสอบความเป็นพหุมิติ และความถูกต้องของโปรแกรมที่ใช้ในการจำลองข้อมูล

สำหรับการวิจัยในครั้งนี้ซึ่งเป็นการวิจัยเชิงทดลองที่ศึกษาจากการจำลองข้อมูลโดยไม่ได้ทำการศึกษากับข้อมูลจริง ดังนั้นหัวใจสำคัญสำหรับการศึกษาโดยใช้การจำลองข้อมูลคือการตรวจสอบความถูกต้องของข้อมูลที่ได้ โดยเฉพาะอย่างยิ่งการตรวจสอบลักษณะ 2 มิติ ตามเงื่อนไขที่ทำการศึกษา และการตรวจสอบความถูกต้องของโปรแกรมที่ใช้ในการจำลองข้อมูลและโปรแกรมคำนวณวิธีการเชื่อมโยงคะแนนในแต่ละวิธี โดยมีรายละเอียดดังนี้

4.1 การตรวจสอบลักษณะ 2 มิติ

สำหรับการศึกษาในครั้งนี้ผู้วิจัยมีการตรวจสอบข้อมูลที่ได้จากการจำลองเบื้องต้นจากการตรวจสอบโดยการวิเคราะห์องค์ประกอบ (Factor Analysis) โดยพิจารณาจากค่าไอเกน (Eigen Value) และน้ำหนักองค์ประกอบ (Factor Loading) ซึ่งเป็นค่าที่ช่วยพิจารณาว่าข้อสอบควรจัดอยู่ในมิติที่ 1 หรือมิติที่ 2 ในแต่ละมิติให้พิจารณาค่าน้ำหนักองค์ประกอบของ

แต่ละข้อ ถ้าค่าน้ำหนักองค์ประกอบใดมีค่ามาก (เข้าสู่ +1 หรือ -1) ควรจัดข้อนั้นอยู่ในมิติดังกล่าว ในบางกรณีค่าน้ำหนักองค์ประกอบ มีค่ากลางๆ เช่น พบว่า น้ำหนักองค์ประกอบ ของข้อที่ 5 ในมิติที่ 1 เป็น 0.41 และในมิติที่ 2 เป็น 0.51 ทำให้ไม่แน่ใจว่าควรจัดข้อนั้นไว้ในมิติที่ 1 หรือมิติที่ 2 ควรทำการหมุนแกนใหม่

อย่างไรก็ตามในการตรวจสอบดังกล่าวไม่เพียงพอต่อการตรวจสอบลักษณะ 2 มิติ เป็นเพียงผลเบื้องต้นก่อนนำไปประมาณค่าพารามิเตอร์ (Calibrate) โดยใช้โปรแกรม TESTFACT การตรวจสอบที่สำคัญต้องพิจารณาจากค่าเฉลี่ยที่เหลือ (Mean Residual) ซึ่งสามารถพิจารณาได้จากโปรแกรม TESTFACT โดยใช้เกณฑ์ให้ค่า Mean Residual น้อยกว่า 0.05 จึงยอมรับว่าโมเดลที่ได้สอดคล้องกับข้อมูลเชิงประจักษ์ ซึ่งถ้าผลการตรวจสอบพบว่ามีค่ามากกว่า 0.05 ผู้วิจัยจะทำการสร้างข้อมูลใหม่

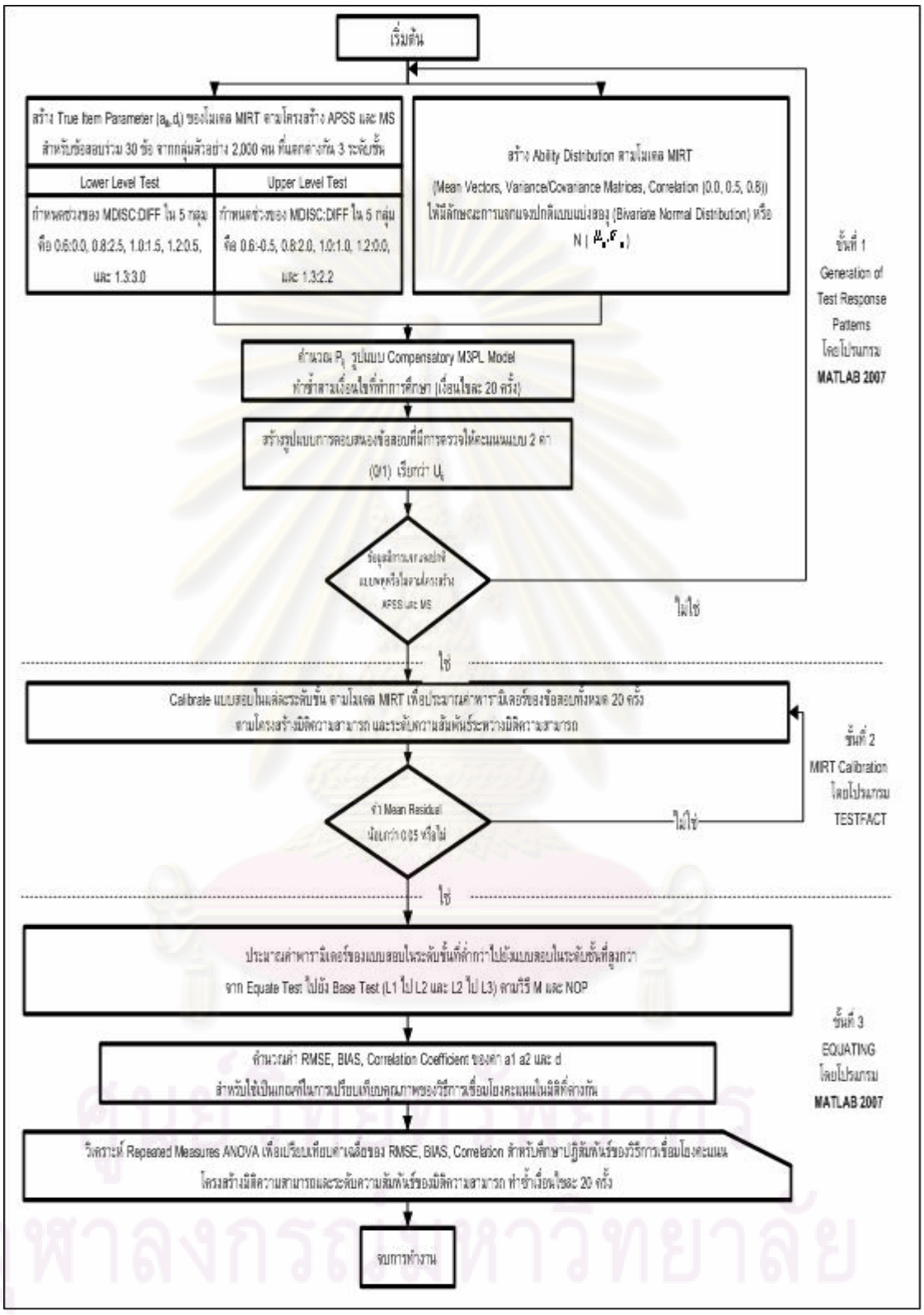
4.2 การตรวจสอบความถูกต้องของโปรแกรมที่ใช้ในการจำลองข้อมูล

ผู้วิจัยตรวจสอบความถูกต้องของโปรแกรมที่พัฒนาขึ้นโดยผู้เชี่ยวชาญที่มีความรู้และเชี่ยวชาญเกี่ยวกับการจำลองข้อมูลโดยใช้โปรแกรม MATLAB ประกอบกับการตรวจสอบ โดยใช้โปรแกรม EXCEL เพื่อตรวจสอบความถูกต้องของค่าที่ได้ ซึ่งโปรแกรมที่พัฒนาขึ้นมีความถูกต้องก็ต่อเมื่อได้ผลตรงกันของทั้งสองโปรแกรม อย่างไรก็ตามในการใช้โปรแกรม EXCEL ในการตรวจสอบ จะใช้ได้เฉพาะการคำนวณสมการเชิงคณิตศาสตร์ เท่านั้น ซึ่งผลการตรวจสอบที่ได้พบว่าทั้งสองโปรแกรมให้ผลการคำนวณเที่เท่ากัน

4.3 การตรวจสอบความถูกต้องของโปรแกรมที่พัฒนาขึ้นสำหรับวิธีการเชื่อมโยงคะแนนในแต่ละวิธี

ผู้วิจัยใช้หลักการตรวจสอบเช่นเดียวกับการตรวจสอบความถูกต้องของโปรแกรมที่ใช้ในการจำลองข้อมูล แต่อย่างไรก็ตามการเชื่อมโยงคะแนนโดยใช้วิธี M ไม่สามารถตรวจสอบโดยใช้โปรแกรม EXCEL ได้ครบทุกขั้นตอน เนื่องจากมีความยุ่งยากซ้อน อย่างไรก็ตามผู้วิจัยมีการตรวจสอบความถูกต้องจากผู้เชี่ยวชาญ ขณะที่วิธี NOP สามารถคำนวณโดยใช้โปรแกรม EXCEL ได้โดยตรง แต่จะต้องระมัดระวังในขั้นตอนของการคำนวณค่าระหว่างเวคเตอร์หรือเมตริกซ์ เพราะจะทำให้ผลที่ได้เกิดข้อผิดพลาดได้ง่าย ซึ่งจากผลการตรวจสอบพบว่าทั้งสองโปรแกรมให้ผลการคำนวณเที่เท่ากัน

การดำเนินงานทั้งหมดสามารถสรุปขั้นตอนการจำลองข้อมูลและการวิเคราะห์ข้อมูลดังแผนภาพที่ 3.4



แผนภาพที่ 3.4 ขั้นตอนการจำลองข้อมูลและการวิเคราะห์ข้อมูล

บทที่ 4
ผลการวิเคราะห์ข้อมูล

การศึกษาในครั้งนี้เป็นการวิจัยเชิงทดลองโดยใช้ข้อมูลจำลอง เพื่อเปรียบเทียบวิธีการเชื่อมโยงคะแนนที่มีลักษณะการหมุนแกนแตกต่างกัน 2 วิธี คือวิธีการหมุนแกนแบบตั้งฉากโดยวิธี M และการหมุนแกนแบบไม่ตั้งฉากโดยวิธี NOP มุ่งเน้นสำหรับการศึกษาภาคตัดขวาง ซึ่งจำเป็นต้องอาศัยกระบวนการสร้างสเกลแนวตั้งตามโมเดล MIRT เป็นการแปลงคะแนนค่าพารามิเตอร์ของข้อสอบจากระดับชั้นหนึ่งไปยังอีกระดับชั้นหนึ่งที่เป็นระดับชั้นติดกันและต่อเนื่องกัน เพื่อให้อยู่ในสเกลที่สามารถเปรียบเทียบกันได้ เมื่อแบบสอบต่างฉบับมุ่งวัดลักษณะเดียวกันแต่ศึกษาใน 2 มิติ โดยมีระดับความยากต่างกันและกลุ่มผู้สอบมีการแจกแจงความสามารถอยู่คนละประชากร และแยกศึกษาในแต่ละมิติ ซึ่งการศึกษาในครั้งนี้มุ่งศึกษา 3 ระดับชั้น คือ ระดับที่ 1-3 กำหนดให้แบบสอบในระดับที่สูงกว่าเป็นแบบสอบฐาน ส่วนแบบสอบในระดับชั้นที่ต่ำกว่ากำหนดให้เป็นแบบสอบปรับเทียบคะแนน โดยการออกแบบครั้งนี้ นักเรียนในแต่ละระดับชั้นจะทำแบบสอบในระดับของตนเองและแบบสอบในระดับที่สูงกว่าที่อยู่ติดกัน ซึ่งการจำลองข้อมูลจะกำหนดให้นักเรียนในระดับที่ 1 ทำแบบสอบระดับที่ 2 (L1_2) สำหรับนักเรียนในระดับที่ 2 ทำแบบสอบในระดับตนเองและทำแบบสอบระดับที่ 3 (L2_3) ส่วนนักเรียนระดับที่ 3 ทำแบบสอบในระดับชั้นของตนเอง

ในการนำเสนอผลการศึกษาจึงมุ่งเน้นการวิเคราะห์คุณภาพของการเชื่อมโยงคะแนนที่ได้จากการแปลงค่าพารามิเตอร์ของข้อสอบจากแบบสอบปรับเทียบคะแนนไปสู่แบบสอบฐานในแต่ละมิติ นั่นคือ ค่าอำนาจจำแนกของข้อสอบในมิติที่ 1 (a_1) ค่าอำนาจจำแนกของข้อสอบในมิติที่ 2 (a_2) และค่าพารามิเตอร์ความยากของข้อสอบ (d) ประกอบด้วย RMSE BIAS และค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าที่แปลงได้จากแบบสอบฐานและค่าที่ได้จากจากการประมาณจากแบบสอบฐาน แทนด้วยสัญลักษณ์ “CORR” โดยมีเกณฑ์ในการพิจารณาแต่ละดัชนี ดังนี้

1) RMSE แสดงถึงความคงที่ (Stability) ของการเชื่อมโยงคะแนน โดยมีหลักเกณฑ์ในการพิจารณาจากการเปรียบเทียบค่าที่ได้ ถ้ามีค่าดัชนี RMSE น้อยกว่า แสดงว่ามีความคงที่ของการเชื่อมโยงคะแนนสูงกว่า สะท้อนให้เห็นถึงความคลาดเคลื่อนของการเชื่อมโยงคะแนนมีน้อย แต่ถ้ามีค่าดัชนี RMSE มากกว่า แสดงว่ามีความคงที่ของการเชื่อมโยงคะแนนต่ำ สะท้อนให้เห็นถึงความคลาดเคลื่อนของการเชื่อมโยงคะแนนมีมาก ซึ่งงานวิจัยครั้งนี้ทำการเปรียบเทียบความแตกต่างระหว่างค่าเฉลี่ยจาก Ln RMSE ซึ่งเป็นค่าที่ผกผันกับ RMSE นั่นคือถ้ามีค่า Ln RMSE มาก จะมีค่า RMSE น้อย แสดงถึงความคงที่ของการเชื่อมโยงคะแนน แต่ถ้ามีค่า Ln RMSE น้อย จะมีค่า RMSE มาก แสดงถึงความคลาดเคลื่อนของการเชื่อมโยงคะแนน

2) BIAS แสดงถึงความถูกต้อง (Accuracy) ของการเชื่อมโยงคะแนน โดยมีหลักเกณฑ์ในการพิจารณาจากการเปรียบเทียบค่าที่ได้ คือ ถ้ามีค่าดัชนี BIAS ใกล้ศูนย์มากกว่าสะท้อนให้เห็นถึงการประมาณค่าใกล้เคียงกับค่าที่แท้จริงมากกว่า โดยเครื่องหมายที่ได้จะเป็นตัวสะท้อนถึงความคลาดเคลื่อนของการประมาณค่าที่ห่างจากค่าที่แท้จริงในลักษณะใด นั่นคือ ถ้าดัชนี BIAS ติดลบ (-) แสดงว่ามีการประมาณค่าต่ำกว่าความเป็นจริง (Underestimate) ถ้าเป็นบวก (+)แสดงว่ามีการประมาณค่าสูงกว่าความเป็นจริง (Overestimate)

3) CORR แสดงถึงความสัมพันธ์ระหว่างค่าที่แปลงได้จากแบบสอบปรับเทียบคะแนนกับคะแนนเกณฑ์ที่ได้จากแบบสอบฐาน โดยมีหลักเกณฑ์ในการพิจารณาจากการเปรียบเทียบค่าที่ได้ ถ้ามีค่า CORR สูง จะสะท้อนให้เห็นถึงคุณภาพของการเชื่อมโยงคะแนนที่ดี

โดยแยกการนำเสนอคุณภาพของการเชื่อมโยงคะแนนเป็น 2 ลักษณะ คือ 1) ผลที่ได้จากการแปลงค่าพารามิเตอร์ของข้อสอบจากระดับที่ 1 ไปยังระดับที่ 2 (L1_2) และ 2) ผลที่ได้การแปลงค่าพารามิเตอร์ของข้อสอบจากระดับที่ 2 ไปยังระดับที่ 3 (L2_3)

ผลการวิเคราะห์ข้อมูลถูกนำเสนอออกเป็น 3 ตอน ได้แก่

ตอนที่ 1 ผลการพิจารณาคุณภาพของการเชื่อมโยงคะแนนในภาพรวมโดยการวิเคราะห์ความแปรปรวนแบบพหุจำแนกตามวิธีการเชื่อมโยงคะแนน โครงสร้างมิติความสามารถและระดับความสัมพันธ์ระหว่างมิติความสามารถ เพื่อศึกษาปฏิสัมพันธ์ของตัวแปรต้น คือ วิธีการเชื่อมโยงคะแนน โครงสร้างมิติความสามารถและระดับความสัมพันธ์ระหว่างมิติความสามารถ พร้อมทั้งเปรียบเทียบรายคู่สำหรับเงื่อนไขที่มีนัยสำคัญทางสถิติที่ระดับ .05 ซึ่งผู้วิจัยได้เสนอค่าสถิติพื้นฐานและกราฟอธิบายภาพปฏิสัมพันธ์ระหว่างตัวแปร โดยผู้วิจัยได้เลือกเงื่อนไขหรือภาพของปฏิสัมพันธ์ที่น่าสนใจ และมีความหมายมากที่สุดมานำเสนอ โดยมุ่งเน้นการเปรียบเทียบระหว่างวิธี M และวิธี NOP ซึ่ง Winer (1971, อ้างใน สุมาลี ชูศรี, 2536) ได้ให้ข้อเสนอแนะไว้ว่าในการอธิบายปฏิสัมพันธ์ระหว่างตัวประกอบหรือตัวแปร 3 ตัว สิ่งที่ผู้อธิบายควรคำนึงถึง คือ การพิจารณาเฉพาะจุดที่น่าสนใจและมีความหมายมากที่สุดเพื่อนำจุดเด่นเหล่านั้นมาตีความหมายให้ชัดเจนยิ่งขึ้น โดยแยกนำเสนอเป็นตอนย่อย ดังนี้

1.1 ผลการพิจารณาคุณภาพของการเชื่อมโยงคะแนนในภาพรวมของการแปลงค่าพารามิเตอร์ข้อสอบจากระดับที่ 1 ไปยังระดับที่ 2 (L1_2)

1.1.1 ค่าสถิติเบื้องต้นของดัชนีบ่งชี้คุณภาพการเชื่อมโยงคะแนน

1.1.2 ผลการวิเคราะห์ความแปรปรวนแบบพหุจำแนก 3 ทาง (Three-way MANOVA)

1.2 ผลการพิจารณาคุณภาพของการเชื่อมโยงคะแนนในภาพรวมของการแปลงค่าพารามิเตอร์ข้อสอบจากระดับที่ 2 ไปยังระดับที่ 3 (L2_3)

1.2.1 ค่าสถิติเบื้องต้นของดัชนีบ่งชี้คุณภาพการเชื่อมโยงคะแนน

1.2.2 ผลการวิเคราะห์ความแปรปรวนแบบพหุจำแนก 3 ทาง (Three-way MANOVA)

ตอนที่ 2 ผลการเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนนรายสถานการณ์ เพื่อเปรียบเทียบความแตกต่างระหว่างวิธีการเชื่อมโยงคะแนน 2 วิธี กับเงื่อนไขที่ศึกษา คือ โครงสร้างมิติความสามารถและระดับความสัมพันธ์ระหว่างมิติความสามารถที่ต่างกัน โดยนำเสนอเป็น 2 ลักษณะคือ 1) ผลการเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนนระหว่างวิธี M กับวิธี NOP ของแต่ละโครงสร้างมิติความสามารถ (APSS และ MS) และแต่ละระดับความสัมพันธ์มิติความสามารถ (r=0.0, r=0.5 และ r=0.8) โดยใช้สถิติทดสอบ One-way MANOVA 2) ผลการเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนนระหว่างวิธี M กับวิธี NOP ภายใต้ 6 เงื่อนไข คือ APSS_0.0, MS_0.0, APSS_0.0, MS_0.5, APSS_0.8 และ MS_0.8 โดยใช้สถิติทดสอบ t พร้อมทั้งนำเสนอแผนภาพในรูป BOX Plot เพื่อศึกษาลักษณะของการกระจายข้อมูลที่มีการทำซ้ำ 20 ครั้ง ทั้ง 6 เงื่อนไข ตามรายดัชนีบ่งชี้คุณภาพทั้ง 3 ชนิด คือ RMSE BIAS และ CORR ทั้งในมิติที่ 1 และ มิติที่ 2 โดยแยกนำเสนอเป็นตอนย่อย ดังนี้

2.1 ผลการเปรียบเทียบค่าเฉลี่ยของดัชนี RMSE BIAS และค่า CORR จากระดับที่ 1 ไปยังระดับที่ 2 รายสถานการณ์

2.2 ผลการเปรียบเทียบค่าเฉลี่ยของดัชนี RMSE BIAS และค่า CORR จากระดับที่ 2 ไปยังระดับที่ 3 รายสถานการณ์

ตอนที่ 3 สรุปผลการวิเคราะห์ตามวัตถุประสงค์การวิจัย ซึ่งประกอบด้วย 2 ข้อ คือ 1) ศึกษาปฏิสัมพันธ์ระหว่างวิธีการเชื่อมโยงคะแนน โครงสร้างมิติความสามารถและระดับความสัมพันธ์ระหว่างมิติความสามารถ และ 2) เปรียบเทียบคุณภาพของการเชื่อมโยงคะแนนของวิธีการเชื่อมโยงคะแนนกับเงื่อนไขที่ศึกษา

ผู้วิจัยได้กำหนดสัญลักษณ์และอักษรย่อภาษาอังกฤษที่ใช้สื่อความหมายแทนชุดข้อมูลเงื่อนไขที่ทำการศึกษา และค่าสถิติ ดังนี้

a_1	แทน	ค่าอำนาจจำแนกของข้อสอบในมิติที่ 1
a_2	แทน	ค่าอำนาจจำแนกของข้อสอบในมิติที่ 2
d	แทน	ค่าความยากของข้อสอบ
M	แทน	วิธีการหมุนแกนแบบตั้งฉากโดยวิธี M
NOP	แทน	วิธีการหมุนแกนแบบไม่ตั้งฉากโดยวิธี NOP
APSS	แทน	โครงสร้างมิติความสามารถที่ไม่ซับซ้อน

MS	แทน	โครงสร้างมิติความสามารถที่ซับซ้อนมากขึ้น
r=0.0	แทน	ระดับความสัมพันธ์ระหว่างมิติความสามารถเท่ากับ 0.0
r=0.5	แทน	ระดับความสัมพันธ์ระหว่างมิติความสามารถเท่ากับ 0.5
r=0.8	แทน	ระดับความสัมพันธ์ระหว่างมิติความสามารถเท่ากับ 0.8
APSS_0.0	แทน	ไม่มีความสัมพันธ์ระหว่างมิติความสามารถที่มีโครงสร้างไม่ซับซ้อน
APSS_0.5	แทน	ความสัมพันธ์ระหว่างมิติความสามารถอยู่ในระดับปานกลางที่มีโครงสร้างไม่ซับซ้อน
APSS_0.8	แทน	ความสัมพันธ์ระหว่างมิติความสามารถอยู่ในระดับสูงที่มีโครงสร้างไม่ซับซ้อน
MS_0.0	แทน	ไม่มีความสัมพันธ์ระหว่างมิติความสามารถที่มีโครงสร้างซับซ้อน
MS_0.5	แทน	ความสัมพันธ์ระหว่างมิติความสามารถอยู่ในระดับปานกลางที่มีโครงสร้างซับซ้อน
MS_0.8	แทน	ความสัมพันธ์ระหว่างมิติความสามารถอยู่ในระดับสูงที่มีโครงสร้างซับซ้อน
RMSE	แทน	ค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย
RMSE_a_1	แทน	ค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ยของค่าอำนาจจำแนกในมิติที่ 1
RMSE_a_2	แทน	ค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ยของค่าอำนาจจำแนกในมิติที่ 2
BIAS	แทน	ค่าความลำเอียง
BIAS_a1	แทน	ค่าความลำเอียงของค่าอำนาจจำแนกในมิติที่ 1
BIAS_a1	แทน	ค่าความลำเอียงของค่าอำนาจจำแนกในมิติที่ 2
CORR	แทน	ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าพารามิเตอร์ที่ประมาณได้จากแบบสอบฐานและค่าพารามิเตอร์ที่ได้จากการแปลงคะแนนจากแบบสอบปรับเทียบคะแนน
CORR_a_1	แทน	ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าพารามิเตอร์ที่ประมาณได้จากแบบสอบฐานและค่าพารามิเตอร์ที่ได้จากการแปลงคะแนนจากแบบสอบปรับเทียบคะแนนของค่าอำนาจจำแนกในมิติที่ 1
CORR_a_2	แทน	ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าพารามิเตอร์ที่ประมาณได้จากแบบสอบฐานและค่าพารามิเตอร์ที่ได้จากการแปลงคะแนนจากแบบสอบปรับเทียบคะแนนของค่าอำนาจจำแนกในมิติที่ 2
L1_2	แทน	การแปลงค่าพารามิเตอร์ของข้อสอบจากระดับที่ 1 ไปยังระดับที่ 2
L2_3	แทน	การแปลงค่าพารามิเตอร์ของข้อสอบจากระดับที่ 2 ไปยังระดับที่ 3

ตอนที่ 1 ผลการพิจารณาคุณภาพของการเชื่อมโยงคะแนนในภาพรวมโดยการวิเคราะห์ความแปรปรวนแบบพหุจำแนกตามวิธีการเชื่อมโยงคะแนน โครงสร้างมิติความสามารถ และระดับความสัมพันธ์ระหว่างมิติความสามารถ

เนื่องจากการศึกษาครั้งนี้มีจุดมุ่งหมายสำคัญ คือ การศึกษาปฏิสัมพันธ์ของวิธีการเชื่อมโยงคะแนน โครงสร้างมิติความสามารถ และระดับความสัมพันธ์ระหว่างมิติความสามารถ ซึ่งเป็นตัวแปรต้นของการวิจัย พร้อมทั้งการเปรียบเทียบความแตกต่างของดัชนีบ่งชี้คุณภาพทั้ง 3 ตัว คือ ดัชนี RMSE BIAS และ CORR ของค่าพารามิเตอร์ข้อสอบทั้งในมิติที่ 1 และมิติที่ 2 ซึ่งเป็นตัวแปรตามของการวิจัย จำแนกตามวิธีการเชื่อมโยงคะแนน โครงสร้างมิติความสามารถ และระดับความสัมพันธ์ระหว่างมิติความสามารถ ดังนั้นสถิติที่เหมาะสมสำหรับการศึกษาตัวแปรตามหลายตัวแปร โดยจำแนกตามตัวแปรต้น 3 ตัวแปร เพื่อให้สามารถตอบคำถามการวิจัยได้ในคราวเดียวกัน คือ การวิเคราะห์ความแปรปรวนแบบพหุจำแนก 3 ทาง (3-WAY MANOVA) โดยทำการพิจารณาในแต่ระดับชั้นของการแปลงคะแนน ได้ผลการวิเคราะห์ ดังนี้

1.1 ผลการพิจารณาคุณภาพของการเชื่อมโยงคะแนนในภาพรวมของการแปลงค่าพารามิเตอร์ข้อสอบจากระดับที่ 1 ไปยังระดับที่ 2 (L1_2)

คุณภาพของวิธีการเชื่อมโยงคะแนนในการวิจัยครั้งนี้ พิจารณาจากดัชนีค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (Root-Mean Square Error; RMSE) ค่าความลำเอียง (BIAS) และค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient; CORR) ของค่าพารามิเตอร์ที่ประมาณได้จากแบบสอบฐานและค่าพารามิเตอร์ที่ได้จากการแปลงคะแนนจากแบบสอบปรับเทียบคะแนน ดังนั้นจึงจำเป็นต้องพิจารณาค่าสถิติเบื้องต้นของดัชนีดังกล่าวเพื่อศึกษาลักษณะของข้อมูลของผลที่ได้จากการแปลงค่าพารามิเตอร์ของข้อสอบจากระดับที่ 1 ไปยังระดับที่ 2 (L1_2) ก่อนทำการวิเคราะห์ความแปรปรวนแบบพหุเพื่อศึกษาปฏิสัมพันธ์ระหว่าง 3 ตัวแปรต่อไป

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

1.1.1 ค่าสถิติเบื้องต้นของดัชนีบ่งชี้คุณภาพการเชื่อมโยงคะแนน

การวิเคราะห์ข้อมูลในเบื้องต้นนี้เป็นการวิเคราะห์ด้วยค่าสถิติพื้นฐานเพื่อบรรยายลักษณะของดัชนี BIAS RMSE และ CORR ข้อมูลในส่วนนี้ประกอบด้วย ค่าต่ำสุด (MIN) ค่าสูงสุด (MAX) ค่าเฉลี่ย (M) และ ส่วนเบี่ยงเบนมาตรฐาน (SD) ของผลที่ได้จากการแปลงค่าพารามิเตอร์ของข้อสอบจากระดับที่ 1 ไประดับที่ 2 เมื่อพิจารณาค่าสถิติพื้นฐานของผลที่ได้จากการแปลงค่าพารามิเตอร์ข้อสอบ พบว่า ดัชนี RMSE BIAS และ CORR ของที่ได้จากการแปลงค่าพารามิเตอร์ความยากของข้อสอบ (d) มีค่าไม่ต่างกันทั้ง วิธี M และวิธี NOP อันเนื่องมาจากผลการแปลงค่าพารามิเตอร์ความยากของข้อสอบ (d) จากแบบสอบปรับเทียบคะแนนไปยังแบบสอบฐานมีค่าเท่ากันทุกครั้งของการทำซ้ำ (20 ครั้ง) ในแต่ละเงื่อนไขของการจำลองข้อมูล แม้จะใช้วิธีการแตกต่างกัน ดังนั้นการเสนอผลครั้งนี้จึงมุ่งเน้นเปรียบเทียบผลที่ได้จากการแปลงค่าพารามิเตอร์อำนาจจำแนกในมิติที่ 1 (a_1) และค่าอำนาจจำแนกในมิติที่ 2 (a_2) เท่านั้น สำหรับผลที่ได้จากการแปลงค่าพารามิเตอร์ความยากของข้อสอบ (d) จะนำเสนอค่าดัชนีบ่งชี้คุณภาพเพื่อใช้สำหรับการพิจารณาระดับความมากน้อยของความคลาดเคลื่อนและความสัมพันธ์ที่เกิดขึ้นเพื่อเป็นสารสนเทศสำคัญในการนำไปเลือกใช้ในสถานการณ์ต่าง ๆ ได้อย่างเหมาะสม โดยมีรายละเอียดดังแสดงในตารางที่ 4.1

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.1 ค่าสถิติเบื้องต้นสำหรับการพิจารณาดัชนี RMSE BIAS และ CORR ของผลที่ได้จากการแปลงค่าพารามิเตอร์ข้อสอบจากระดับที่ 1 ไปยังระดับที่ 2

สถิติ	เงื่อนไขการจำลองข้อมูล			RMSE			BIAS			CORR		
				a1	a2	d	a1	a2	d	a1	a2	d
MIN	r= 0.0	APSS	M	0.094	0.111	0.205	-0.149	-0.149	-0.006	0.850	0.870	0.896
			NOP	0.093	0.109	0.205	-0.142	-0.142	-0.006	0.866	0.865	0.896
		MS	M	0.121	0.096	0.125	-0.205	-0.214	0.070	0.716	0.723	0.682
			NOP	0.120	0.093	0.125	-0.218	-0.196	0.070	0.773	0.704	0.682
	r= 0.5	APSS	M	0.090	0.087	0.138	-0.116	-0.135	0.304	0.705	0.773	0.942
			NOP	0.090	0.087	0.138	-0.093	-0.112	0.304	0.684	0.765	0.942
		MS	M	0.108	0.094	0.114	-0.154	-0.151	0.427	0.659	0.713	0.959
			NOP	0.107	0.086	0.114	-0.121	-0.109	0.427	0.638	0.706	0.959
	r= 0.8	APSS	M	0.115	0.106	0.135	-0.130	-0.098	0.375	0.428	0.575	0.951
			NOP	0.098	0.096	0.135	-0.105	-0.079	0.375	0.384	0.461	0.951
		MS	M	0.144	0.142	0.110	-0.166	-0.271	-0.682	0.267	0.224	0.867
			NOP	0.131	0.110	0.110	-0.122	-0.206	-0.682	0.295	0.332	0.867
MAX	r= 0.0	APSS	M	0.231	0.183	0.337	-0.018	-0.069	0.240	0.960	0.949	0.956
			NOP	0.221	0.182	0.337	-0.045	-0.075	0.240	0.960	0.951	0.956
		MS	M	0.262	0.197	0.557	-0.107	-0.102	0.425	0.914	0.925	0.982
			NOP	0.243	0.193	0.557	-0.051	-0.062	0.425	0.913	0.921	0.982
	r= 0.5	APSS	M	0.203	0.153	0.266	-0.051	-0.067	0.399	0.919	0.936	0.977
			NOP	0.195	0.150	0.266	-0.050	-0.061	0.399	0.919	0.936	0.977
		MS	M	0.199	0.156	0.218	-0.082	-0.070	0.495	0.872	0.889	0.988
			NOP	0.172	0.143	0.218	0.002	-0.017	0.495	0.872	0.892	0.988
	r= 0.8	APSS	M	0.225	0.198	0.303	-0.002	-0.012	0.509	0.832	0.827	0.984
			NOP	0.206	0.179	0.303	0.010	0.002	0.509	0.827	0.819	0.984
		MS	M	0.254	0.262	0.422	0.009	0.053	0.510	0.721	0.712	0.987
			NOP	0.232	0.229	0.422	0.051	0.108	0.510	0.745	0.760	0.987
M	r= 0.0	APSS	M	0.157	0.143	0.271	-0.103	-0.121	0.156	0.906	0.917	0.932
			NOP	0.154	0.142	0.271	-0.115	-0.118	0.156	0.914	0.918	0.932
		MS	M	0.151	0.157	0.253	-0.166	-0.144	0.311	0.862	0.822	0.933
			NOP	0.147	0.154	0.253	-0.128	-0.117	0.311	0.865	0.818	0.933
	r= 0.5	APSS	M	0.135	0.120	0.213	-0.080	-0.091	0.366	0.858	0.872	0.963
			NOP	0.133	0.116	0.213	-0.077	-0.081	0.366	0.857	0.865	0.963
		MS	M	0.134	0.130	0.181	-0.115	-0.115	0.461	0.801	0.785	0.972
			NOP	0.129	0.122	0.181	-0.062	-0.073	0.461	0.797	0.782	0.972
	r= 0.8	APSS	M	0.157	0.147	0.196	-0.059	-0.048	0.414	0.678	0.721	0.969
			NOP	0.148	0.134	0.196	-0.046	-0.033	0.414	0.659	0.710	0.969
		MS	M	0.194	0.204	0.187	-0.091	-0.076	0.367	0.499	0.495	0.966
			NOP	0.160	0.163	0.187	-0.023	-0.016	0.367	0.589	0.588	0.966
SD	r= 0.0	APSS	M	0.033	0.022	0.040	0.033	0.023	0.061	0.027	0.020	0.015
			NOP	0.029	0.021	0.040	0.028	0.021	0.061	0.025	0.021	0.015
		MS	M	0.038	0.028	0.084	0.021	0.027	0.071	0.053	0.054	0.061
			NOP	0.034	0.028	0.084	0.050	0.041	0.071	0.042	0.059	0.061
	r= 0.5	APSS	M	0.027	0.019	0.034	0.017	0.016	0.023	0.051	0.047	0.011
			NOP	0.026	0.018	0.034	0.011	0.012	0.023	0.055	0.051	0.011
		MS	M	0.020	0.017	0.023	0.021	0.018	0.019	0.054	0.051	0.006
			NOP	0.015	0.018	0.023	0.043	0.028	0.019	0.056	0.053	0.006
	r= 0.8	APSS	M	0.032	0.026	0.042	0.025	0.025	0.034	0.103	0.075	0.010
			NOP	0.031	0.025	0.042	0.024	0.023	0.034	0.111	0.098	0.010
		MS	M	0.025	0.032	0.066	0.042	0.066	0.278	0.129	0.144	0.026
			NOP	0.026	0.028	0.066	0.052	0.069	0.278	0.113	0.119	0.026

จากตารางที่ 4.1 *เมื่อพิจารณาเงื่อนไขไม่มีความสัมพันธ์ระหว่างมิติสำหรับโครงสร้างมิติความสามารถที่ไม่ซับซ้อน (APSS_0.0)* ของผลที่ได้จากการแปลงค่าพารามิเตอร์ของข้อสอบระดับที่ 1 ไป 2 โดยพิจารณาจากค่าต่ำสุดและสูงสุด พบว่า วิธี M จะให้ค่าต่ำสุด ($RMSE\_a_1$=0.094, $RMSE\_a_2$=0.111, $BIAS\_a_1$=-0.149, $BIAS\_a_2$=-0.149) มีคลาดเคลื่อนสูงกว่าและมีการประมาณค่าที่ต่ำกว่าความเป็นจริง (Underestimate) มากกว่าวิธี NOP ($RMSE\_a_1$=0.093, $RMSE\_a_2$=0.109, $BIAS\_a_1$=-0.142, $BIAS\_a_2$=-0.142) อย่างเห็นได้ชัด เนื่องจากมีค่าดัชนี BIAS ติดลบ เมื่อพิจารณาค่าสูงสุด พบว่าวิธี NOP มีความคงที่ของการประมาณค่ามากกว่าวิธี M ($RMSE\_a_1$=0.221, $RMSE\_a_2$=0.182) แต่มีการประมาณค่าที่ต่ำกว่าความเป็นจริงมากกว่า ($BIAS\_a_1$=-0.045, $BIAS\_a_2$=-0.075) นอกจากนี้เมื่อพิจารณาค่าเฉลี่ยของดัชนี RMSE พบว่าวิธี M ($RMSE\_a_1$=0.157, $RMSE\_a_2$=0.143) และวิธี NOP ($RMSE\_a_1$=0.154, $RMSE\_a_2$=0.142) มีค่าใกล้เคียงกันทั้ง 2 วิธี โดยวิธี M มีความคลาดเคลื่อนสูงกว่าเพียงเล็กน้อย สำหรับดัชนี BIAS พบว่า วิธี NOP ($BIAS\_a_1$=-0.115, $BIAS\_a_2$=-0.118) มีความคลาดเคลื่อนมากกว่าวิธี M ($BIAS\_a_1$=-0.103, $BIAS\_a_2$=-0.121) โดยมีการประมาณค่าต่ำกว่าความเป็นจริงอย่างเห็นได้ชัด สำหรับค่าต่ำสุดของ CORR พบว่า ค่าพารามิเตอร์อำนาจจำแนกในมิติที่ 1 ของวิธี NOP มีค่าสัมประสิทธิ์สหสัมพันธ์สูงกว่าวิธี M ในขณะที่ค่าพารามิเตอร์อำนาจจำแนกในมิติที่ 2 มีค่าน้อยกว่าวิธี M อย่างไรก็ตามค่าที่ได้จากทั้ง 2 วิธีต่างกันเพียงเล็กน้อย

เมื่อพิจารณาเงื่อนไขไม่มีความสัมพันธ์ระหว่างมิติสำหรับโครงสร้างที่ซับซ้อนมากขึ้น (MS_0.0) พบว่า มีลักษณะใกล้เคียงกับโครงสร้างที่ไม่ซับซ้อน แต่ต่างกันในประเด็นของค่าสูงสุดและต่ำสุดของ BIAS โดยวิธี M มีการประมาณค่าต่ำกว่าความเป็นจริง (Underestimate) มากกว่าวิธี NOP ในทุกค่าพารามิเตอร์อำนาจจำแนก และค่าสูงสุดของ CORR พบว่า วิธี M ($CORR\_a_1$=0.914, $CORR\_a_1$=0.925) มีค่าสูงกว่าวิธี NOP ($CORR\_a_1$=0.913, $CORR\_a_1$=0.921) ทุกค่าพารามิเตอร์อำนาจจำแนก แสดงว่า ค่าพารามิเตอร์ที่ได้จากการแปลงคะแนนโดยวิธี M มีความสัมพันธ์กับคะแนนเกณฑ์มากกว่าวิธี NOP เมื่อพิจารณาจากค่าสูงสุด นั่นเอง

เมื่อพิจารณาเงื่อนไขความสัมพันธ์ระหว่างมิติในระดับปานกลางสำหรับโครงสร้างไม่ซับซ้อน (APSS_0.5) โดยพิจารณาจากค่าต่ำสุดและสูงสุด พบว่า พิสัยของดัชนี RMSE และ BIAS ในวิธี M กว้างกว่าวิธี NOP ในทุกค่าพารามิเตอร์ แสดงว่าวิธี M จะให้ความคลาดเคลื่อนสูงกว่าวิธี NOP อย่างไรก็ตามทั้ง 2 มีค่า RMSE ต่ำสุดไม่ต่างกัน ($RMSE\_a_1$=0.090, $RMSE\_a_2$=0.087) แต่วิธี M มีค่าสูงสุดของ RMSE ($RMSE\_a_1$=0.203, $RMSE\_a_2$=0.153) สูงกว่า

วิธี NOP (RMSE_a_1=0.195, RMSE_a_2=0.150) แสดงว่าวิธี M แสดงช่วงของความคลาดเคลื่อนสูงกว่าวิธี NOP รวมทั้งวิธี M มีค่าต่ำสุด (BIAS_a_1=-0.116, BIAS_a_2=-0.135) และค่าสูงสุด (BIAS_a_1=-0.051, BIAS_a_2=-0.067) ของ BIAS ต่ำกว่าความเป็นจริงมากกว่าวิธี NOP ซึ่งสอดคล้องกับค่าเฉลี่ยของ RMSE ที่พบว่าวิธี M (RMSE_a_1=0.135, RMSE_a_2=0.120) มีความคลาดเคลื่อนสูงกว่าวิธี NOP (RMSE_a_1=0.133, RMSE_a_2=0.116) และมีค่าเฉลี่ยของ BIAS ต่ำกว่าความเป็นจริงมากกว่า NOP สำหรับการพิจารณาค่า CORR พบว่า วิธี M มีค่าต่ำสุดของสัมประสิทธิ์สหสัมพันธ์สูงกว่าวิธี NOP ในขณะที่ทั้งสองวิธีให้ค่าสูงสุดของ CORR ไม่แตกต่างกัน เมื่อพิจารณาค่าเฉลี่ยของสัมประสิทธิ์สหสัมพันธ์ของทั้ง 2 วิธี พบว่าวิธี NOP มีค่าสูงกว่า วิธี M แสดงว่า ค่าพารามิเตอร์ที่ได้จากการแปลงคะแนนโดยวิธี NOP มีความสัมพันธ์กับคะแนนเกณฑ์มากกว่าวิธี M

เมื่อพิจารณาเงื่อนไขความสัมพันธ์ระหว่างมิติในระดับปานกลางสำหรับโครงสร้างที่ซับซ้อนมากขึ้น (MS_0.5) พบว่า วิธี M มีค่าต่ำสุดของ RMSE และ BIAS (RMSE_a_1=0.108, RMSE_a_2=0.094, BIAS_a_1=-0.154, BIAS_a_2=-0.151) สูงกว่า วิธี NOP (RMSE_a_1=0.107, RMSE_a_2=0.086, BIAS_a_1=-0.121, BIAS_a_2=-0.109) ในทุกค่าพารามิเตอร์ เช่นเดียวกับค่าสูงสุดที่ให้ผลเช่นเดียวกัน แสดงว่า วิธี M ให้ค่าความคลาดเคลื่อนของค่าต่ำสุดและสูงสุดมากกว่าวิธี NOP รวมถึงมีการประมาณค่าที่ต่ำกว่าความเป็นจริงมากกว่า โดยมีข้อสังเกตว่าวิธี NOP ให้ค่าสูงสุดของ BIAS (BIAS_a1=0.002) ที่มากกว่าความเป็นจริงเล็กน้อยเนื่องจากมีค่าเป็นบวก เมื่อพิจารณาค่าเฉลี่ยของ RMSE พบว่าวิธี M (RMSE_a_1=0.134, RMSE_a_2=0.130, BIAS_a_1=-0.115, BIAS_a_2=-0.115) มีความคลาดเคลื่อนสูงและมีค่าเฉลี่ยของ BIAS ต่ำกว่าความเป็นจริงมากกว่าวิธี NOP (RMSE_a_1=0.129, RMSE_a_2=0.122, BIAS_a_1=-0.062, BIAS_a_2=-0.073) สำหรับค่าต่ำสุดและสูงสุดของสัมประสิทธิ์สหสัมพันธ์ พบว่า วิธี M มีค่าสูงกว่าวิธี NOP ในขณะที่ทั้งสองวิธีให้ค่าสูงสุดของสัมประสิทธิ์สหสัมพันธ์ไม่แตกต่างกันสำหรับค่าพารามิเตอร์ a_2 แต่วิธี NOP ให้ค่าสัมประสิทธิ์สหสัมพันธ์มากกว่าสำหรับค่าพารามิเตอร์ a_1 แต่เมื่อพิจารณาค่าเฉลี่ยของสัมประสิทธิ์สหสัมพันธ์ของทั้ง 2 วิธี พบว่าวิธี NOP มีค่าสูงกว่า วิธี M แสดงว่าค่าพารามิเตอร์ที่ได้จากการแปลงคะแนนโดยวิธี NOP มีความสัมพันธ์กับคะแนนเกณฑ์มากกว่าวิธี M

เมื่อพิจารณาเงื่อนไขความสัมพันธ์ระหว่างมิติในระดับสูงสำหรับโครงสร้างที่ไม่ซับซ้อน (APSS_0.8) พบว่า มีลักษณะใกล้เคียงกับเงื่อนไข MS_0.5 แต่ต่างกันในประเด็นของค่าต่ำสุด สูงสุดและค่าเฉลี่ยของสัมประสิทธิ์สหสัมพันธ์ โดยพบว่าวิธี M มีค่าสูงกว่า วิธี NOP ในทุกค่าพารามิเตอร์ แสดงว่าวิธี NOP มีความคงที่ของการประมาณค่าพาราเตอร์และมีการ

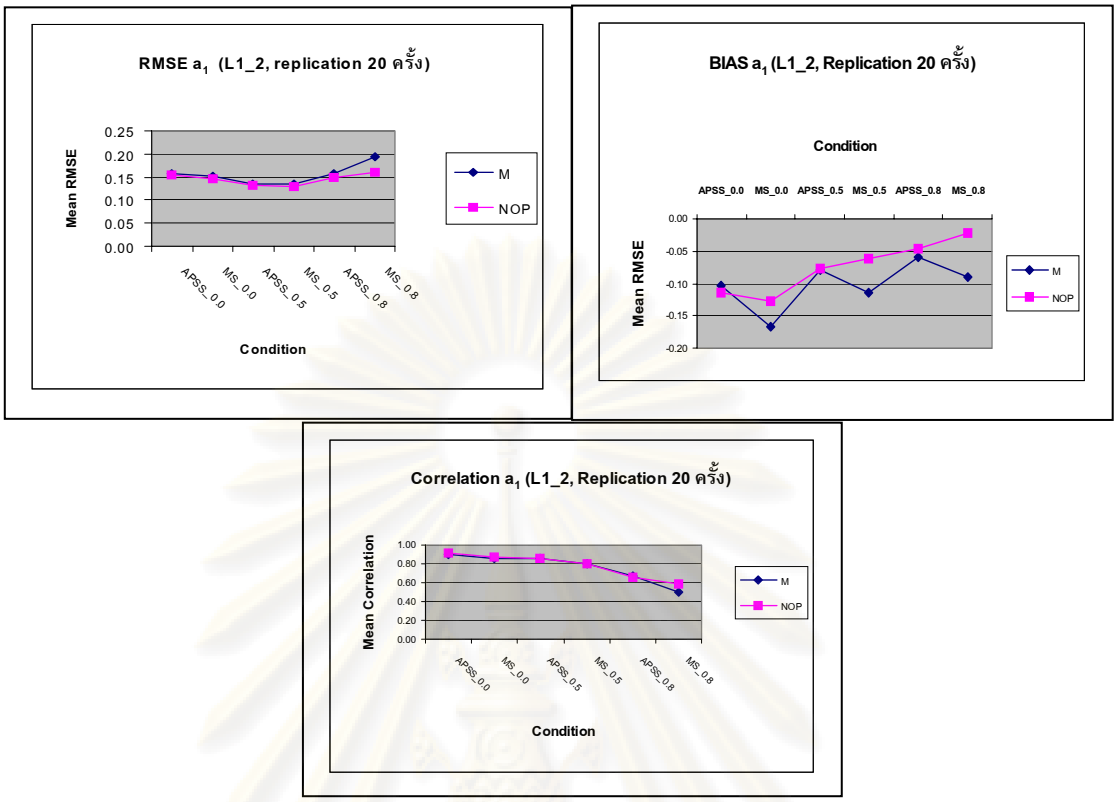
ประมาณค่าพารามิเตอร์ใกล้เคียงกับค่าที่แท้จริงมากกว่าวิธี M อีกทั้งวิธี NOP มีค่าความสัมพันธ์ของคะแนนที่แปลงได้กับคะแนนเกณฑ์สูงกว่าวิธี M

เมื่อพิจารณาเงื่อนไขความสัมพันธ์ระหว่างมิติในระดับสูงสำหรับโครงสร้างที่ซับซ้อนมากขึ้น (MS_0.8) พบว่า มีลักษณะที่คล้ายคลึงกับเงื่อนไขอื่น นั่นคือ วิธี M มีความคลาดเคลื่อนสูงกว่าวิธี NOP อย่างเห็นได้ชัดในทุกค่าพารามิเตอร์ แสดงให้เห็นว่าวิธี NOP มีความคงที่ของการประมาณค่าพาราเตอร์และมีการประมาณค่าพารามิเตอร์ใกล้เคียงกับค่าที่แท้จริงมากกว่าวิธี M อีกทั้งวิธี NOP มีค่าความสัมพันธ์ของคะแนนที่แปลงได้กับคะแนนเกณฑ์สูงกว่าวิธี M แต่ต่างจากเงื่อนไขอื่นในประเด็นของค่า BIAS สูงสุด ของวิธี M ($BIAS\_a_1$=0.009, $BIAS\_a_2$=0.053) และวิธี NOP ($BIAS\_a_1$=0.051, $BIAS\_a_2$=0.108) มีแนวโน้มที่จะประมาณค่าสูงกว่าความเป็นจริง แต่วิธี M มีค่าสูงกว่าความเป็นจริงเพียงเล็กน้อยโดยมีค่าข้าใกล้ศูนย์มากกว่าวิธี NOP

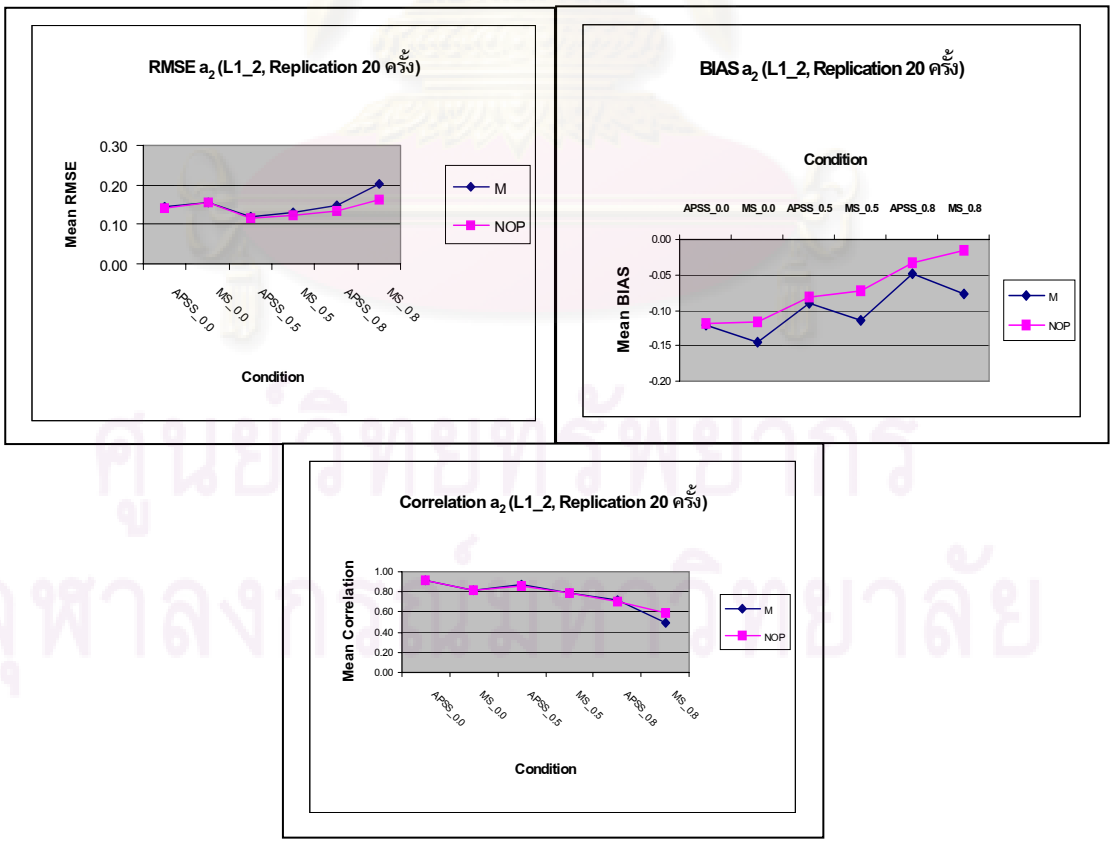
เมื่อพิจารณาค่าพารามิเตอร์ d โดยภาพรวม พบว่า เมื่อความสัมพันธ์ระหว่างมิติความสามารถสูงขึ้น ทั้ง 2 วิธีมีผลทำให้เกิดความคลาดเคลื่อนของการประมาณค่าลดลง แต่มีแนวโน้มในการประมาณค่าสูงกว่าความเป็นจริงสูงขึ้น โดยเมื่อไม่มีความสัมพันธ์หรือความสัมพันธ์ระหว่างมิติความสามารถต่ำ มีแนวโน้มที่จะประมาณค่าต่ำกว่าความเป็นจริง ซึ่งช่วงพิสัยของค่า RMSE ในโครงสร้างที่ซับซ้อนจะน้อยกว่าโครงสร้างที่ไม่ซับซ้อน ในขณะที่โครงสร้างที่ซับซ้อนมีแนวโน้มที่จะประมาณค่าสูงกว่าความเป็นจริงมากกว่าโครงสร้างที่ไม่ซับซ้อน สำหรับค่าสัมประสิทธิ์สหสัมพันธ์พบว่ามีค่าใกล้เคียงกันคือมีค่าเข้าใกล้ 1 อย่างไรก็ตามพบว่าเมื่อความสัมพันธ์ระหว่างมิติความสามารถอยู่ในระดับสูงมีผลทำให้มีค่าสัมประสิทธิ์สหสัมพันธ์เพิ่มมากขึ้น

จากที่กล่าวมาเป็นการบรรยายถึงค่าสถิติเบื้องต้นของดัชนี RMSE BIAS และ CORR ทั้ง 6 เงื่อนไข คือ APSS_0.0, MS_0.0, APSS_0.0, MS_0.5, APSS_0.8 และ MS_0.8 โดยมุ่งเน้นการเปรียบเทียบระหว่างวิธี M และ NOP เพื่อให้ง่ายต่อความเข้าใจสามารถสรุปรวมทั้ง 6 เงื่อนไข ในแต่ละค่าพารามิเตอร์ของการแปลงคะแนน ดังแผนภาพต่อไปนี้

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



แผนภาพที่ 4.1 ค่าเฉลี่ยดัชนี RMSE BIAS และ CORR ของ a_1 ในการทำซ้ำ 20 ครั้ง ของ 6 เงื่อนไขที่ศึกษา เปรียบเทียบระหว่างวิธี M และ NOP

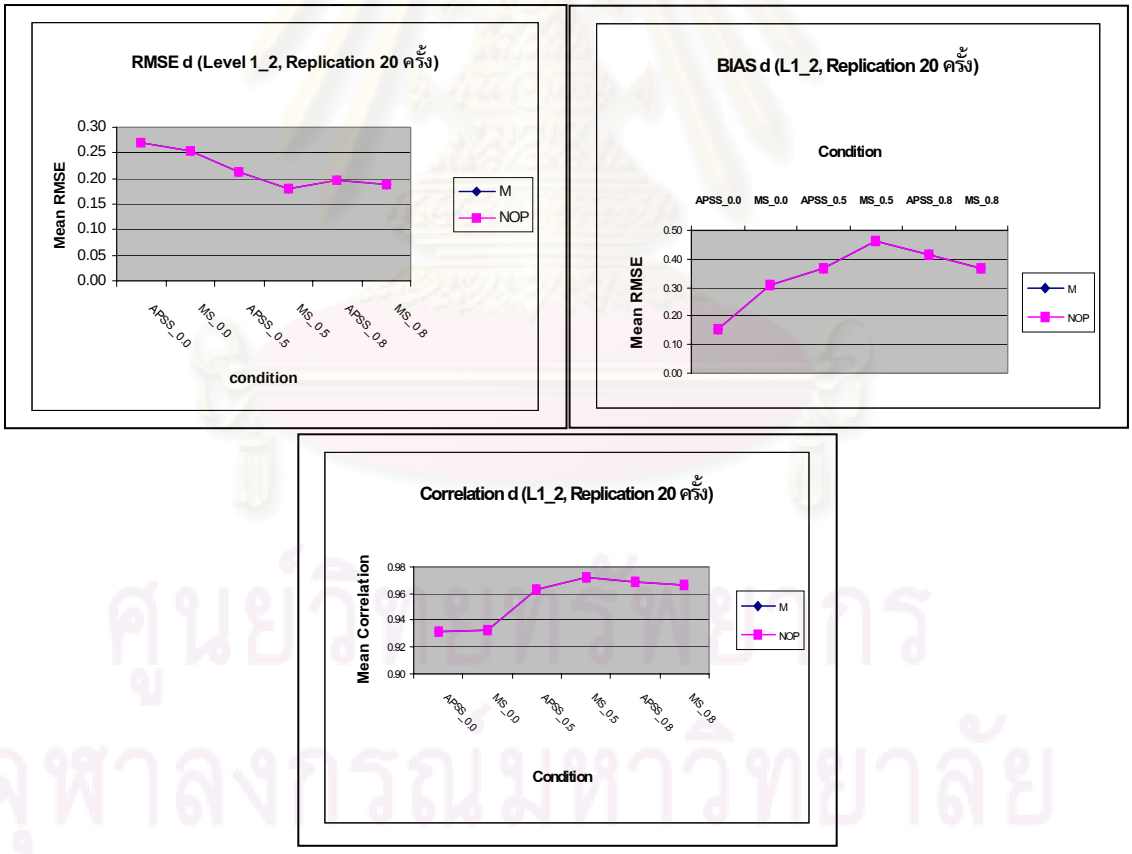


แผนภาพที่ 4.2 ค่าเฉลี่ยดัชนี RMSE BIAS และ CORR ของ a_2 ในการทำซ้ำ 20 ครั้ง ของ 6 เงื่อนไขที่ศึกษา เปรียบเทียบระหว่างวิธี M และ NOP

จากแผนภาพที่ 4.1 และ 4.2 ให้ผลการแปลงคะแนนของค่าพารามิเตอร์อำนาจจำแนกในมิติที่ 1 และ 2 ใกล้เคียงกัน นั่นคือ เมื่อพิจารณาจากดัชนี RMSE พบว่า การแปลงค่าพารามิเตอร์ในมิติที่ 1 มีความคงที่ของการแปลงคะแนนระหว่างวิธี M และวิธี NOP ใกล้เคียงกันเมื่อศึกษากับเงื่อนไขที่ไม่มีระดับความสัมพันธ์หรือความสัมพันธ์ในระดับกลาง แต่เมื่อความสัมพันธ์ระหว่างมิติความสามารถสูงขึ้นวิธี NOP จะมีความคงที่มากกว่าวิธี M อย่างเห็นได้ชัด

เมื่อพิจารณาจากดัชนี BIAS พบว่าทั้งสองวิธี มีการประมาณค่าที่ต่ำกว่าความเป็นจริง เนื่องจากมีค่าดัชนีติดลบ อย่างไรก็ตามวิธี NOP มีการประมาณค่าที่ใกล้เคียงกับค่าที่แท้จริงมากกว่าวิธี M อย่างเห็นได้ชัด โดยโครงสร้างที่ไม่ซับซ้อน (APSS) จะมีการประมาณค่าใกล้เคียงกับค่าที่แท้จริงมากกว่าโครงสร้างที่ซับซ้อนมากขึ้น (MS)

เมื่อพิจารณาจากค่า CORR พบว่า ทั้งสองวิธีมีค่าใกล้เคียงกันมากและมีค่าเข้าใกล้ 1 แต่เมื่อความสัมพันธ์ระหว่างมิติความสามารถสูง พบว่า วิธี NOP มีค่าสูงกว่าวิธี M แสดงว่าวิธี NOP มีค่าความสัมพันธ์ของคะแนนที่แปลงได้กับคะแนนเกณฑ์สูงกว่าวิธี M



แผนภาพที่ 4.3 ค่าเฉลี่ยดัชนี RMSE BIAS และ CORR ของ d ในการทำซ้ำ 20 ครั้ง ของ 6 เงื่อนไขที่ศึกษาเปรียบเทียบระหว่างวิธี M และ NOP

จากแผนภาพที่ 4.3 เมื่อพิจารณาการแปลงค่าพารามิเตอร์ความยาก (d) พบว่า มีค่าดัชนี RMSE BIAS และ CORR เท่ากันทั้งวิธี M และวิธี NOP โดยเมื่อพิจารณา RMSE พบว่าเมื่อความสัมพันธ์ระหว่างมิติเพิ่มขึ้นและมีโครงสร้างซับซ้อนมากขึ้น พบว่ามีความคงที่ของการแปลงคะแนนมากกว่า ซึ่งเมื่อเปรียบเทียบระหว่างโครงสร้างที่ซับซ้อนกับไม่ซับซ้อน พบว่าโครงสร้างที่ซับซ้อนมีความคงที่ของการเชื่อมโยงคะแนนมากกว่าโครงสร้างที่ไม่ซับซ้อน โดยจากแผนภาพพบว่าเงื่อนไข MS_0.5 มีความคงที่ของการแปลงคะแนนมากกว่าเงื่อนไขอื่น

เมื่อพิจารณาดัชนี BIAS พบว่า เมื่อความสัมพันธ์ระหว่างมิติเพิ่มขึ้นและมีโครงสร้างมีการประมาณค่าที่ใกล้เคียงกับค่าที่แท้จริงของการแปลงคะแนนมากกว่า ซึ่งเมื่อเปรียบเทียบระหว่างโครงสร้างที่ซับซ้อนกับไม่ซับซ้อน พบว่าโครงสร้างที่ไม่ซับซ้อนมีการประมาณค่าใกล้เคียงกับค่าที่แท้จริงมากกว่าโครงสร้างที่ซับซ้อน โดยจากแผนภาพพบว่าเงื่อนไข APSS_0.0 มีความถูกต้องของการแปลงคะแนนมากกว่าเงื่อนไขอื่น

เมื่อพิจารณา CORR พบว่า เมื่อความสัมพันธ์ระหว่างมิติความสามารถเพิ่มขึ้นค่าความสัมพันธ์ระหว่างค่าที่ได้จากการแปลงคะแนนกับคะแนนเกณฑ์มีค่าสูงขึ้น แต่อย่างไรก็ตามพบว่าทั้ง 6 เงื่อนไขที่ศึกษามีค่าเข้าใกล้ 1

โดยสรุป พบว่า วิธี M และ วิธี NOP มีค่าดัชนีบ่งชี้คุณภาพใกล้เคียงกัน ในเงื่อนไขไม่มีความสัมพันธ์หรือความสัมพันธ์ระหว่างมิติความสามารถปานกลาง แต่เมื่อความสัมพันธ์ระหว่างมิติเพิ่มขึ้นและโครงสร้างเชิงมิติมีความซับซ้อนมากขึ้น พบว่า วิธี NOP มีความคงที่และมีความถูกต้องในการประมาณค่ามากกว่าวิธี M โดยวิธี M แนวโน้มในการประมาณค่าที่ต่ำกว่าความแป็นจริง

เมื่อพิจารณาถึงความสัมพันธ์ระหว่างค่าที่แปลงได้จากแบบสอบปรับเทียบคะแนนกับคะแนนเกณฑ์ พบว่ามีค่าความสัมพันธ์สูงทั้งสองวิธี โดยมีค่าเข้าใกล้ 1 อย่างไรก็ตามเมื่อความสัมพันธ์ระหว่างมิติความสามารถเพิ่มขึ้นพบว่า NOP มีค่าความสัมพันธ์ของคะแนนที่แปลงได้กับคะแนนเกณฑ์สูงกว่าวิธี M

จากที่กล่าวมาข้างต้นเป็นการบรรยายค่าสถิติเบื้องต้นของดัชนีบ่งชี้คุณภาพของการเชื่อมโยงคะแนน ทำให้สามารถพิจารณาข้อมูลพื้นฐานสำคัญในการนำไปสู่การทดสอบสมมติฐานทางสถิติในการศึกษาปฏิสัมพันธ์ระหว่างวิธีการเชื่อมโยงคะแนน โครงสร้างมิติความสามารถ และระดับความสัมพันธ์ระหว่างมิติความสามารถ ซึ่งส่งผลต่อคุณภาพของการเชื่อมโยงคะแนน โดยการวิเคราะห์ความแปรปรวนแบบพหุ (MANOVA) ดังที่จะกล่าวถึงในรายละเอียดต่อไป

1.1.2 ผลการวิเคราะห์ความแปรปรวนแบบพหุจำแนก 3 ทาง (Three-way MANOVA)

การนำเสนอผลการวิเคราะห์ข้อมูลในส่วนนี้ เป็นการนำเสนอผลการวิเคราะห์ความแปรปรวนแบบจำแนก 3 ทาง (Three-way MANOVA) เพื่อศึกษาปฏิสัมพันธ์ระหว่างวิธีการเชื่อมโยงคะแนน โครงสร้างมิติความสามารถและระดับความสัมพันธ์ระหว่างมิติความสามารถ ซึ่งส่งผลต่อคุณภาพของการเชื่อมโยงคะแนนที่พิจารณาจากค่าดัชนี RMSE BIAS และ CORR ของการแปลงค่าพารามิเตอร์อำนาจจำแนกในมิติที่ 1 (a_1) และมิติที่ 2 (a_2) ส่วนค่าพารามิเตอร์ความยาก (d) ไม่ได้นำมาวิเคราะห์เนื่องจากผลที่ได้จากการแปลงคะแนนระหว่างวิธี M และ วิธี NOP ไม่แตกต่างกัน โดยที่แต่ละตัวแปรต้น แบ่งเป็นหลายระดับคือ 1) วิธีการเชื่อมโยงคะแนน แบ่งเป็น 2 ระดับ คือ วิธี M และ วิธี NOP 2) โครงสร้างมิติความสามารถแบ่งเป็น 2 ระดับ คือ โครงสร้างที่ไม่ซับซ้อน (APSS) และโครงสร้างที่ซับซ้อนมากขึ้น (MS) และ 3) ความสัมพันธ์ระหว่างมิติความสามารถแบ่งเป็น 3 ระดับ คือ ไม่มีระดับความสัมพันธ์ (r= 0.0) ความสัมพันธ์ระดับกลาง (r=0.5) และความสัมพันธ์ระดับสูง (r=0.8)

ผู้วิจัยวิเคราะห์ข้อมูลในส่วนนี้เพื่อตรวจสอบความแตกต่างของเซ็นทรอยด์ ซึ่งเป็นความแตกต่างระหว่างค่าเฉลี่ยของของตัวแปรดัชนีบ่งชี้คุณภาพ RMSE BIAS และ CORR ของค่าพารามิเตอร์อำนาจจำแนกในมิติที่ 1 และ มิติที่ 2 ซึ่งเป็นตัวแปรตามทั้งหมด 6 ตัวแปร คือ 1) ค่าดัชนี Ln RMSE ในมิติที่ 1 (Ln RMSE_a1) 2) ค่าดัชนี Ln RMSE ในมิติที่ 2 (Ln RMSE_a2) 3) ค่าดัชนี BIAS ในมิติที่ 1 (BIAS_a1) 4) ค่าดัชนี BIAS ในมิติที่ 2 (BIAS_a2) 5) ค่า CORR ในมิติที่ 1 (CORR_a_1) และ 6) ค่า CORR ในมิติที่ 2 (CORR_a_2) ได้ผลการวิเคราะห์เพื่อตรวจสอบข้อตกลงเบื้องต้น ดังนี้

ตารางที่ 4.2 การทดสอบ Box's Test และ Bartlett's Test

Box's M	1115.671	Bartlett's Test of Sphericity	
F	4.637	Likelihood Ratio	0.000
df1	210.000	Approx. Chi-Square	1447.353
df2	38576.136	df	20.000
p	0.000	p	0.000

จากตารางที่ 4.3 ผลการตรวจสอบข้อตกลงเบื้องต้นของสถิติวิเคราะห์พบว่า เมตริกซ์ความแปรปรวน-ความแปรปรวนร่วมของตัวแปรทั้ง 6 ตัวแปร คือ Ln RMSE_a1 Ln RMSE_a2 BIAS_a1 BIAS_a2 CORR_a_1 และ CORR_a_2 ต่างกันระหว่างกลุ่มอย่างมีนัยสำคัญทางสถิติ (Box's M =1115.671,df=(210,38576.14), p=.000) ซึ่งเป็นการละเมิดข้อตกลงเบื้องต้นทางสถิติของการวิเคราะห์ความแปรปรวนของตัวแปรพหุนาม ที่กำหนดให้

เมตริกซ์ความแปรปรวน-ความแปรปรวนร่วมของทุกกลุ่มต้องเท่ากัน ซึ่งการทดสอบด้วยวิธี Box's M ค่อนข้างมีความไวต่อการละเมิดข้อตกลงเบื้องต้นดังกล่าว นักวิจัยส่วนใหญ่จึงทดสอบข้อตกลงเบื้องต้นที่ระดับนัยสำคัญทางสถิติ .001 อย่างไรก็ตามจากงานวิจัยของ Holloway & Dunn (1967) Hakstain, Roed & Linn (1979) และ Olson (1974) พบว่า การที่เมตริกซ์ค่าแปรปรวนร่วมไม่เท่ากันจะไม่มีผลกระทบต่อระดับนัยสำคัญในแต่ละกลุ่มย่อยเท่ากัน โดยในการวิจัยครั้งนี้ได้กำหนดให้ขนาดตัวอย่างในแต่ละกลุ่มย่อยเท่ากัน ซึ่งเรียกว่า Balanced Three-way MANOVA (อ้างใน กัลยา วานิชย์บัญชา, 2550) ซึ่งการทดสอบความแปรปรวนพหุด้วย F-test มีความแกร่งเพียงพอ เมื่อมีการละเมิดข้อตกลงเบื้องต้นดังกล่าว (Huynh & Mandeville, 1979 อ้างใน ณัฏฐภรณ์ หลาวทอง และ สกล ซื่อธนาพรกุล, 2550)

อย่างไรก็ดี การตรวจสอบความสัมพันธ์ระหว่างดัชนีบ่งชี้คุณภาพทั้ง 6 ตัวแปรว่ามีความสัมพันธ์แตกต่างจากเมทริกซ์เอกลักษณ์หรือไม่ ด้วยการทดสอบ Bartlett's test พบว่า ทั้ง 6 ตัวแปรมีความสัมพันธ์กัน (Likelihood=.000, Approx Chi-square=1447.353, df=20, p=.000) ด้วยเหตุนี้จึงสามารถวิเคราะห์ความแปรปรวนตัวแปรพหุนามได้

สถิติทดสอบที่ใช้ในการทดสอบความแปรปรวนตัวแปรพหุนาม 4 ชนิด ผู้วิจัยมีการเลือกนำไปใช้ (อ้างใน ณัฏฐภรณ์ หลาวทอง และ สกล ซื่อธนาพรกุล, 2550) ดังนี้

1) Pillai's Trace เป็นสถิติทดสอบโดยใช้หลักการวิเคราะห์จำแนกพหุ (Multiple Discriminant Analysis; MDA) ซึ่งเป็นส่วนหนึ่งของการวิเคอราะห์ MANOVA ในการวิเคราะห์ด้วยสถิติทดสอบอื่นๆ นอกจากนี้ Pillai's Trace ยังทดสอบได้ดีในกรณีข้อมูลไม่มีการกระจายแบบโค้งปกติ

2) สถิติทดสอบ Wilks' Lamda เป็นสถิติทั่วไปที่นิยมใช้ในการทดสอบ ซึ่งมักถูกนำมาใช้ในการคำนวณเมื่อตัวแปรอิสระมีจำนวนกลุ่มมากกว่า 2 กลุ่มขึ้นไป

3) สถิติทดสอบ Hotelling's Trace เป็นสถิติทดสอบทั่วไป มักถูกนำมาใช้ในการคำนวณเมื่อตัวแปรอิสระมีจำนวนกลุ่มเท่ากับ 2 กลุ่ม

4) สถิติทดสอบ Roy's Largest Root เป็นสถิติทดสอบที่มีความแกร่งน้อยที่สุดเมื่อเทียบกับสถิติทดสอบสามลำดับข้างต้น โดยเฉพาะอย่างยิ่งในกรณีที่มีการละเมิดข้อตกลงเบื้องต้นของการวิเคราะห์ความแปรปรวนพหุ นอกจากนี้สถิติทดสอบ Roy's Largest Root ยังเหมาะที่จะใช้เมื่อตัวแปรมีความสัมพันธ์กันในลักษณะเชิงสันตรงเท่านั้น

สำหรับงานวิจัยครั้งนี้จะใช้สถิติทดสอบทั้ง 4 ชนิดเป็นหลักฐานในการทดสอบ อย่างไรก็ตามผู้วิจัยเน้นที่สถิติทดสอบ 2 ชนิด ได้แก่ Pillai's Trace และ Wilks' Lamda เนื่องจากเหมาะสมสำหรับนำมาใช้ในงานวิจัยนี้มากที่สุด ซึ่งมีตัวแปรอิสระจำนวน 3 กลุ่ม

ตารางที่ 4.3 การวิเคราะห์ความแปรปรวนพหุนามดัชนีบ่งชี้คุณภาพของการเชื่อมโยงคะแนนทั้ง 6 ตัวแปร ระหว่างวิธีการเชื่อมโยงคะแนน โครงสร้างมิติความสามารถและระดับความสัมพันธ์ระหว่างมิติความสามารถที่ต่างกัน

Effect		Value	F	Hypothesis df	Error df	p
Intercept	Pillai's Trace	0.997	12289.249	6	223	.000
	Wilks' Lambda	0.003	12289.249	6	223	.000
	Hotelling's Trace	330.652	12289.249	6	223	.000
	Roy's Largest Root	330.652	12289.249	6	223	.000
วิธีการเชื่อมโยงคะแนน (METHOD)	Pillai's Trace	0.302	16.087	6	223	.000
	Wilks' Lambda	0.698	16.087	6	223	.000
	Hotelling's Trace	0.433	16.087	6	223	.000
	Roy's Largest Root	0.433	16.087	6	223	.000
โครงสร้างมิติความสามารถ (DS)	Pillai's Trace	0.582	51.779	6	223	.000
	Wilks' Lambda	0.418	51.779	6	223	.000
	Hotelling's Trace	1.393	51.779	6	223	.000
	Roy's Largest Root	1.393	51.779	6	223	.000
ระดับความสัมพันธ์ระหว่างมิติความสามารถ (r)	Pillai's Trace	1.307	70.468	12	448	.000
	Wilks' Lambda	0.048	132.472	12	446	.000
	Hotelling's Trace	12.429	229.943	12	444	.000
	Roy's Largest Root	11.802	440.610	6	224	.000
METHOD X DS	Pillai's Trace	0.178	8.033	6	223	.000
	Wilks' Lambda	0.822	8.033	6	223	.000
	Hotelling's Trace	0.216	8.033	6	223	.000
	Roy's Largest Root	0.216	8.033	6	223	.000
METHOD X r	Pillai's Trace	0.096	1.881	12	448	.035
	Wilks' Lambda	0.905	1.896	12	446	.033
	Hotelling's Trace	0.103	1.910	12	444	.031
	Roy's Largest Root	0.088	3.274	6	224	.004
DS X r	Pillai's Trace	0.204	4.236	12	448	.000
	Wilks' Lambda	0.803	4.308	12	446	.000
	Hotelling's Trace	0.237	4.381	12	444	.000
	Roy's Largest Root	0.193	7.194	6	224	.000
M X DS X r	Pillai's Trace	0.075	1.457	12	448	.137
	Wilks' Lambda	0.925	1.476	12	446	.130
	Hotelling's Trace	0.081	1.495	12	444	.122
	Roy's Largest Root	0.078	2.918	6	224	.009

การทดสอบความแปรปรวนพหุนามของดัชนี RMSE BIAS และ CORR ของค่าอำนาจจำแนกในมิติที่ 1 และ มิติที่ 2 ทั้ง 6 ตัวแปร ที่มีวิธีการเชื่อมโยงคะแนน โครงสร้างมิติความสามารถและระดับความสัมพันธ์ระหว่างมิติความสามารถต่างกัน ผลการวิเคราะห์ พบว่าไม่มีปฏิสัมพันธ์ระหว่าง 3 ตัวแปร คือ วิธีการเชื่อมโยงคะแนน โครงสร้างมิติความสามารถและระดับความสัมพันธ์ระหว่างมิติความสามารถที่ทำให้ดัชนี RMSE BIAS และ CORR ของค่าอำนาจจำแนกในมิติที่ 1 และ มิติที่ 2 ทั้ง 6 ตัวแปร ต่างกัน แต่เมื่อพิจารณาปฏิสัมพันธ์ระหว่าง 2 ตัวแปร พบว่ามีปฏิสัมพันธ์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ในทุกประเภทของสถิติทดสอบที่ศึกษา แสดงว่ามีปฏิสัมพันธ์ระหว่างวิธีการเชื่อมโยงคะแนนและระดับความสัมพันธ์ระหว่างมิติความสามารถ มีปฏิสัมพันธ์ระหว่างวิธีการเชื่อมโยงคะแนนและโครงสร้างมิติความสามารถ และมีปฏิสัมพันธ์ระหว่างระดับความสัมพันธ์ระหว่างมิติความสามารถและโครงสร้างมิติความสามารถ ซึ่งส่งผลทำให้ดัชนี Ln RMSE BIAS และ CORR ของค่าอำนาจจำแนกในมิติที่ 1 และมิติที่ 2 ทั้ง 6 ตัวแปร โดยเฉลี่ยต่างกัน

กล่าวโดยสรุป วิธีการเชื่อมโยงคะแนน มีคุณภาพแตกต่างกันขึ้นอยู่กับ 1 โครงสร้างมิติความสามารถและ 2) ระดับความสัมพันธ์ระหว่างมิติความสามารถ

นอกจากนี้เมื่อพิจารณาอิทธิพลหลัก พบว่า วิธีการเชื่อมโยงคะแนน ระดับความสัมพันธ์ระหว่างมิติความสามารถ และโครงสร้างมิติความสามารถที่ต่างกันทำให้มีค่าดัชนี Ln RMSE BIAS และ CORR ของค่าอำนาจจำแนกในมิติที่ 1 และมิติที่ 2 โดยเฉลี่ยต่างกัน

เมื่อพิจารณาค่าสถิติจาก Levene's Test เพื่อพิจารณาความแปรปรวนทั้ง 6 ตัวแปร ปรากฏดังตารางที่ 4.4

ตารางที่ 4.4 การทดสอบความแปรปรวนของค่าดัชนี RMSE BIAS และ CORR ของค่าอำนาจจำแนกในมิติที่ 1 และ มิติที่ 2 ทั้ง 6 ตัวแปร จาก Levene's Test

ตัวแปรตาม	Levene's Test			
	F	df1	df2	p
Ln RMSE_a1	1.505	11	228	.131
Ln RMSE_a2	.508	11	228	.897
BIAS_a1	9.245	11	228	.000
BIAS_a2	6.512	11	228	.000
CORR_a_1	7.095	11	228	.000
CORR_a_2	10.467	11	228	.000

จากตารางที่ 4.4 การทดสอบความเท่ากันของความแปรปรวนด้วยวิธี Levene's Test เพื่อตรวจสอบความแปรปรวน ทั้ง 6 ตัวแปร จำแนกตามวิธีการเชื่อมโยงคะแนน ความสัมพันธ์ระหว่างมิติความสามารถ และระดับความสัมพันธ์ระหว่างมิติความสามารถ พบว่า ความแปรปรวนของ Ln RMSE_a1 (F=1.505, p=.131) และ Ln RMSE_a2 (F=.508, p=.897) จำแนกตามวิธีการเชื่อมโยงคะแนน ความสัมพันธ์ระหว่างมิติความสามารถ และระดับความสัมพันธ์ระหว่างมิติความสามารถ แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ดังนั้นผู้วิจัยจึงเลือกใช้การทดสอบภายหลัง (Post Hoc Tests) หรือการเปรียบเทียบพหุ (Multiple Comparison) เพื่อเปรียบเทียบความแตกต่างของค่าเฉลี่ยแต่ละกลุ่มที่เหมาะสมในกรณีที่ความแปรปรวนของตัวแปรตามในแต่ละกลุ่มมีค่าเท่ากันด้วย Sheffe ในขณะที่ความแปรปรวนของ $BIAS\_a_1$ (F=9.245, p=.000), $BIAS\_a_2$ (F=6.512, p=.000), $CORR\_a_1$ (F=7.095, p=.000) และ $CORR\_a_2$ (F=10.467, p=.000) แตกต่างกันอย่างมีนัยสำคัญทางสถิติ เมื่อใช้วิธีการเชื่อมโยงคะแนน ความสัมพันธ์ระหว่างมิติความสามารถ และระดับความสัมพันธ์ระหว่างมิติความสามารถต่างกัน ดังนั้นผู้วิจัยจึงเลือกใช้การทดสอบภายหลัง (Post Hoc Tests) หรือการเปรียบเทียบพหุ (Multiple Comparison) เพื่อเปรียบเทียบความแตกต่างของค่าเฉลี่ยแต่ละกลุ่มที่เหมาะสมในกรณีที่ความแปรปรวนของตัวแปรตามในแต่ละกลุ่มมีต่างกันด้วย Dunnett T3

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.5 การทดสอบอิทธิพลระหว่างตัวแปรอิสระ (Test of Between-subject Effects)

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	p	การเปรียบเทียบรายคู่
Corrected Model	1. Ln RMSE_a1	2.745	11	0.250	7.455	.000	
	2. Ln RMSE_a2	5.437	11	0.494	17.744	.000	
	3. BIAS_a1	0.339	11	0.031	28.045	.000	
	4. BIAS_a2	0.337	11	0.031	24.166	.000	
	5. CORR_a1	4.015	11	0.365	62.471	.000	
	6. CORR_a2	3.656	11	0.332	59.037	.000	
Intercept	1. Ln RMSE_a1	884.630	1	884.630	26423.945	.000	
	2. Ln RMSE_a2	922.952	1	922.952	33135.864	.000	
	3. BIAS_a1	1.924	1	1.924	1752.303	.000	
	4. BIAS_a2	1.777	1	1.777	1402.198	.000	
	5. CORR_a1	143.684	1	143.684	24591.443	.000	
	6. CORR_a2	143.873	1	143.873	25555.029	.000	
วิธีการเชื่อมโยงคะแนน (METHOD)	1. Ln RMSE_a1	0.176	1	0.176	5.249	.023	M>NOP*
	2. Ln RMSE_a2	0.318	1	0.318	11.414	.001	M>NOP*
	3. BIAS_a1	0.050	1	0.050	45.643	.000	M>NOP**
	4. BIAS_a2	0.041	1	0.041	32.003	.000	M>NOP**
	5. CORR_a1	0.010	1	0.010	1.757	.186	NOP>M
	6. CORR_a2	0.008	1	0.008	1.380	.241	NOP>M
โครงสร้างมิติความสามารถ (DS)	1. Ln RMSE_a1	0.071	1	0.071	2.116	.147	MS>APSS
	2. Ln RMSE_a2	1.154	1	1.154	41.429	.000	MS>APSS**
	3. BIAS_a1	0.015	1	0.015	13.356	.000	MS>APSS**
	4. BIAS_a2	0.004	1	0.004	3.248	.073	MS>APSS
	5. CORR_a1	0.351	1	0.351	60.085	.000	APSS>MS**
	6. CORR_a2	0.848	1	0.848	150.687	.000	APSS>MS**
ระดับความสัมพันธ์ระหว่างมิติความสามารถ (r)	1. Ln RMSE_a1	1.776	2	0.888	26.523	.000	0.8>0.0>0.5**
	2. Ln RMSE_a2	3.135	2	1.567	56.275	.000	0.8>0.0>0.5**
	3. BIAS_a1	0.216	2	0.108	98.223	.000	0.0>0.5>0.8**
	4. BIAS_a2	0.269	2	0.134	106.046	.000	0.0>0.5>0.8**
	5. CORR_a1	3.507	2	1.753	300.075	.000	0.0>0.5>0.8**
	6. CORR_a2	2.626	2	1.313	233.218	.000	0.0>0.5>0.8**
METHOD * DS	1. Ln RMSE_a1	0.046	1	0.046	1.383	.241	
	2. Ln RMSE_a2	0.062	1	0.062	2.226	.137	
	3. BIAS_a1	0.034	1	0.034	31.303	.000	
	4. BIAS_a2	0.017	1	0.017	13.602	.000	
	5. CORR_a1	0.017	1	0.017	2.944	.088	
	6. CORR_a2	0.018	1	0.018	3.172	.076	
METHOD * r	1. Ln RMSE_a1	0.156	2	0.078	2.325	.100	
	2. Ln RMSE_a2	0.229	2	0.114	4.108	.018	
	3. BIAS_a1	0.008	2	0.004	3.848	.023	
	4. BIAS_a2	0.005	2	0.002	1.965	.142	
	5. CORR_a1	0.016	2	0.008	1.386	.252	
	6. CORR_a2	0.026	2	0.013	2.346	.098	

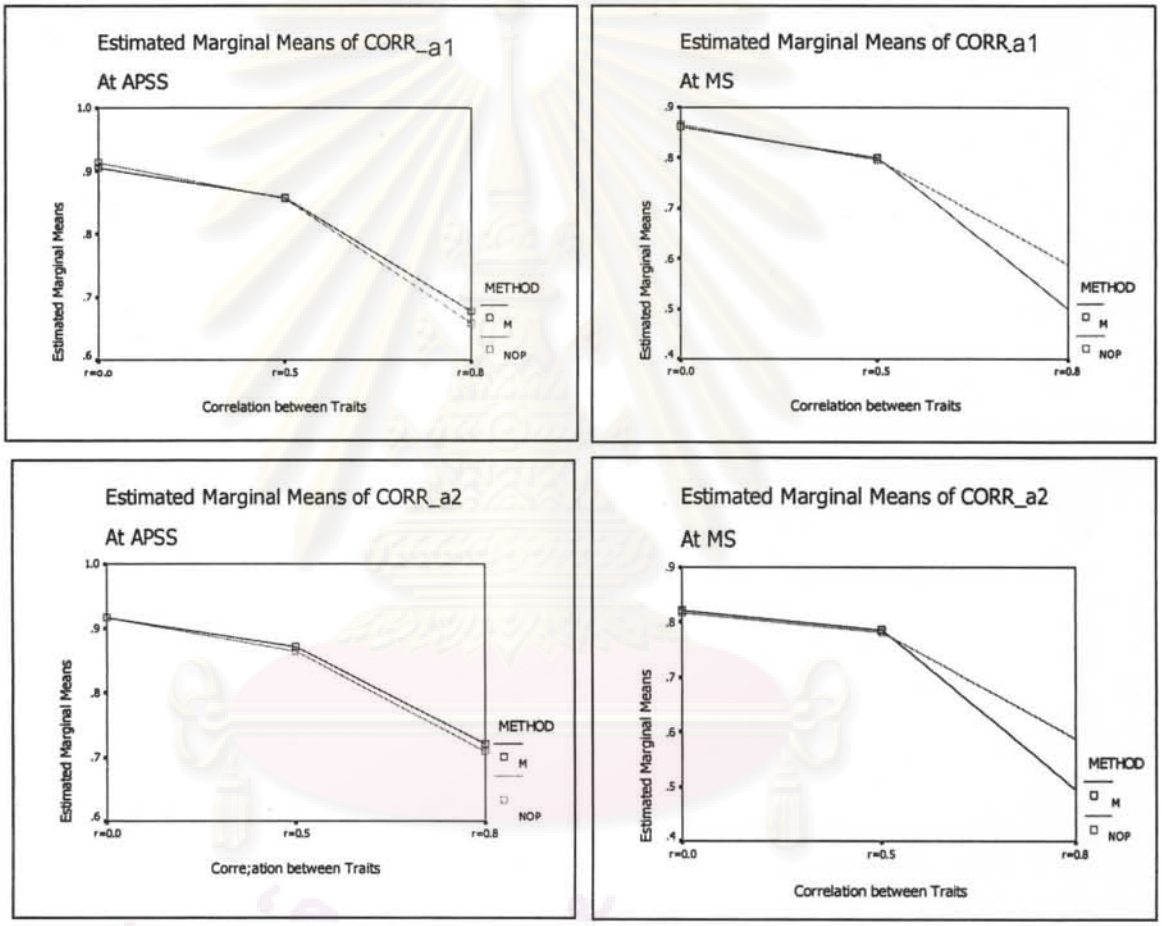
Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	p	การเปรียบเทียบรายคู่
DS * r	1. Ln RMSE_a1	0.470	2	0.235	7.014	.001	
	2. Ln RMSE_a2	0.496	2	0.248	8.899	.000	
	3. BIAS_a1	0.015	2	0.007	6.669	.002	
	4. BIAS_a2	0.000	2	0.000	0.103	.902	
	5. CORR_a1	0.072	2	0.036	6.163	.002	
	6. CORR_a2	0.093	2	0.047	8.266	.000	
METHOD * DS * r	1. Ln RMSE_a1	0.051	2	0.026	0.766	.466	
	2. Ln RMSE_a2	0.043	2	0.022	0.776	.461	
	3. BIAS_a1	0.001	2	0.000	0.356	.701	
	4. BIAS_a2	0.001	2	0.000	0.374	.688	
	5. CORR_a1	0.042	2	0.021	3.573	.030	
	6. CORR_a2	0.037	2	0.018	3.257	.040	
Error	1. Ln RMSE_a1	7.633	228	0.033			
	2. Ln RMSE_a2	6.351	228	0.028			
	3. BIAS_a1	0.250	228	0.001			
	4. BIAS_a2	0.289	228	0.001			
	5. CORR_a1	1.332	228	0.006			
	6. CORR_a2	1.284	228	0.006			
Total	1. Ln RMSE_a1	895.008	240				
	2. Ln RMSE_a2	934.739	240				
	3. BIAS_a1	2.513	240				
	4. BIAS_a2	2.403	240				
	5. CORR_a1	149.032	240				
	6. CORR_a2	148.813	240				
Corrected Total	1. Ln RMSE_a1	10.378	239				
	2. Ln RMSE_a2	11.787	239				
	3. BIAS_a1	0.589	239				
	4. BIAS_a2	0.626	239				
	5. CORR_a1	5.347	239				
	6. CORR_a2	4.940	239				

จากตารางที่ 4.5 สามารถสรุปผลการวิเคราะห์ความแปรปรวนพหุนาม ดังต่อไปนี้

1) ปฏิสัมพันธ์ระหว่าง 3 ตัวแปร (M X DS X r)

ผลการวิเคราะห์ความแปรปรวน พบว่า เซ็นทรอยด์ของค่าสัมประสิทธิ์สหสัมพันธ์ของค่าพารามิเตอร์อำนาจจำแนกในมิติที่ 1 และมิติที่ 2 คือ CORR_a1 ($F_{(2,228)}$=.356 p=.701) และ CORR_a2 ($F_{(2,228)}$=.356 p=.701) มีปฏิสัมพันธ์ระหว่างวิธีการเชื่อมโยงคะแนน (M) โครงสร้างมิติความสามารถ (DS) และระดับความสัมพันธ์ระหว่างมิติความสามารถ (r) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ในขณะที่เซ็นทรอยด์ของดัชนีบ่งชี้

คุณภาพ Ln RMSE BIAS ทั้งอำนาจจำแนกในมิติที่ 1 และมิติที่ 2 คือ RMAE_a1 ($F_{(2,228)}$=.766 p=.466) RMAE_a2 ($F_{(2,228)}$=.776 p=.461) BIAS_a1 ($F_{(2,228)}$= 3.573 p=.030) และ BIAS_a2 ($F_{(2,228)}$= 3.257p=.040) มีปฏิสัมพันธ์อย่างไม่มีนัยสำคัญทางสถิติ นั่นคือ วิธีการเชื่อมโยงคะแนน โครงสร้างมิติความสามารรถ และระดับความสัมพันธ์ระหว่างมิติความสามารถ ไม่มีปฏิสัมพันธ์ที่ส่งผลต่อดัชนีบ่งชี้คุณภาพ RMSE และ BIAS แต่มีปฏิสัมพันธ์กันที่ส่งผลต่อค่า CORR นั่นเอง ซึ่งสามารถแสดงแผนภาพปฏิสัมพันธ์ของวิธีการเชื่อมโยงคะแนนและระดับความสัมพันธ์ระหว่างมิติความสามารถที่ระดับโครงสร้างมิติความสามารถต่างกัน ดังต่อไปนี้



แผนภาพที่ 4.4 กราฟแสดงปฏิสัมพันธ์ระหว่างวิธีการเชื่อมโยงคะแนน ระดับความสัมพันธ์ระหว่างมิติความสามารถในโครงสร้างที่ไม่ซับซ้อน (APSS) และโครงสร้างที่ซับซ้อนมากขึ้น (MS) ซึ่งมีผลต่อค่า CORR (METHOD X r at DS)

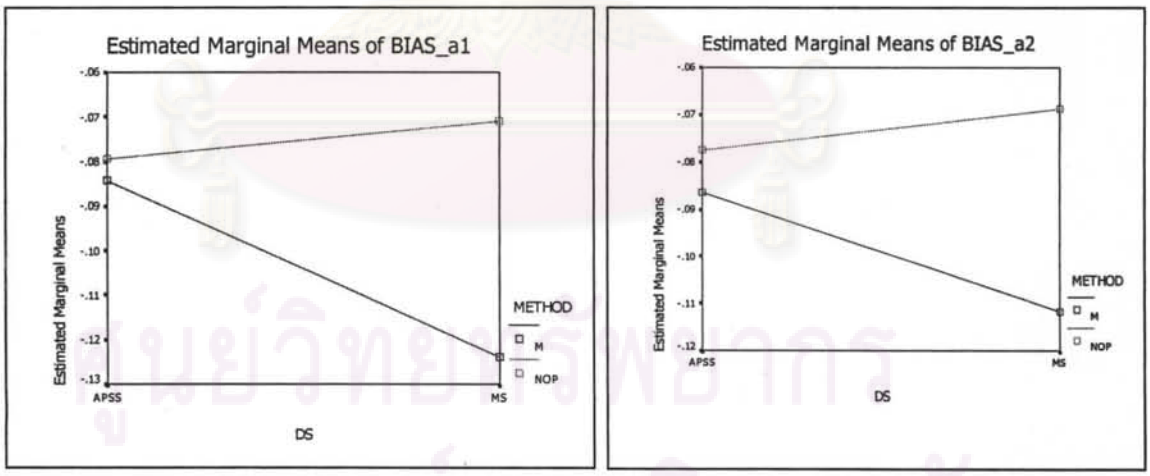
จากแผนภาพแสดงให้เห็นได้อย่างชัดเจนว่า ในกรณีที่แบบสอบมีโครงสร้างที่ไม่ซับซ้อน (APSS) ผลของวิธีการเชื่อมโยงคะแนนและระดับความสัมพันธ์ระหว่างมิติความสามารถมีปฏิสัมพันธ์ร่วมกันในการส่งผลต่อค่า CORR ของค่าอำนาจจำแนกทั้ง 2 มิติ โดยเมื่อความสัมพันธ์ระหว่างมิติความสามารถเท่ากับ 0 และ 0.5 พบว่าวิธี M และวิธี NOP มี

ค่าCORR ใกล้เคียงกัน แต่เมื่อระดับความสัมพันธ์ห่างมิติความสามารถเท่ากับ 0.8 พบว่า วิธี M จะมีค่า CORR มากกว่าวิธี NOP

เมื่อพิจารณาในกรณีที่แบบสอบมีโครงสร้างที่ซับซ้อนมากขึ้น (MS) พบว่า ผลของวิธีการเชื่อมโยงคะแนนและระดับความสัมพันธ์ระหว่างมิติความสามารถมีปฏิสัมพันธ์ร่วมกันในการส่งผลต่อค่า CORR ของค่าอำนาจจำแนกทั้ง 2 มิติ โดยเมื่อความสัมพันธ์ระหว่างมิติความสามารถเท่ากับ 0 และ 0.5 พบว่าวิธี M และวิธี NOP มีค่า CORR ใกล้เคียงกัน เหมือนโครงสร้างที่ไม่ซับซ้อน (APSS) แต่ระดับความสัมพันธ์ห่างมิติความสามารถเท่ากับ 0.8 พบว่า วิธี NOP จะมีค่า CORR มากกว่าวิธี M

2) ปฏิสัมพันธ์ระหว่าง 2 ตัวแปร

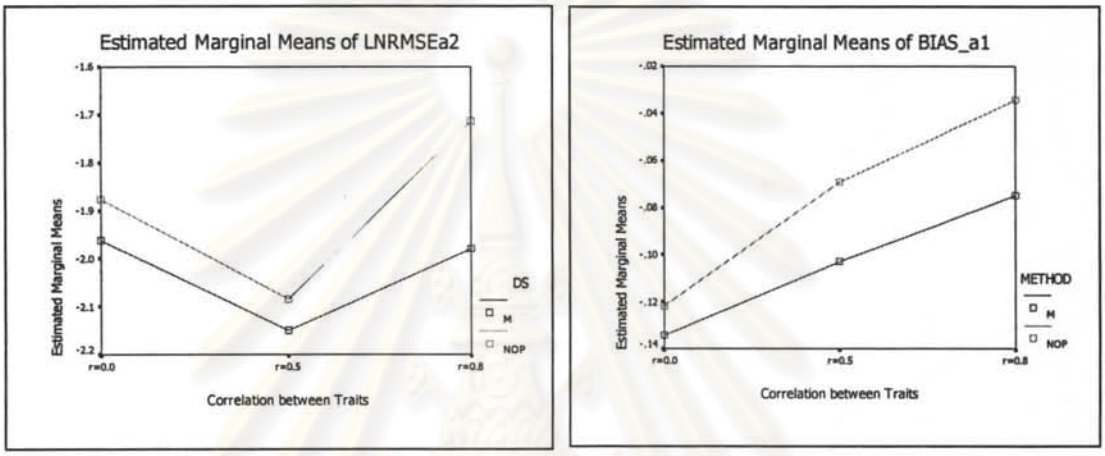
2.1) ปฏิสัมพันธ์ระหว่างวิธีการเชื่อมโยงคะแนนและโครงสร้างมิติความสามารถ (METHOD X DS) พบว่า วิธีการเชื่อมโยงคะแนนและโครงสร้างมิติความสามารถมีปฏิสัมพันธ์ต่อค่าดัชนี BIAS ทั้งในมิติที่ 1 ($F_{(1,228)}$= 31.303, p=.000) และมิติที่ 2 ($F_{(1,228)}$= 13.602, p=.000) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ในขณะไม่มีปฏิสัมพันธ์ที่ส่งผลต่อค่าดัชนี RMSE และ ค่าCORR ทั้งในมิติที่ 1 และ มิติที่ 2 แสดงว่า วิธีการเชื่อมโยงคะแนน และโครงสร้างมิติความสามารถ ไม่มีปฏิสัมพันธ์ที่ส่งผลต่อดัชนีบ่งชี้คุณภาพ RMSE และ CORR แต่มีปฏิสัมพันธ์กันซึ่งส่งผลต่อค่า BIAS นั่นเอง ซึ่งสามารถแสดงแผนภาพปฏิสัมพันธ์ระหว่าง 2 ตัวแปร ดังต่อไปนี้



แผนภาพที่ 4.5 กราฟแสดงปฏิสัมพันธ์ METHOD X DS ในมิติที่ 1 และ 2

จากแผนภาพ จะเห็นว่า เส้นกราฟของวิธี M และ วิธี NOP ไม่ขนานกัน จากภาพไม่ว่าจะเป็นโครงสร้างในลักษณะใดวิธี NOP จะมีค่า BIAS ที่ต่ำกว่าวิธี M แสดงว่าวิธี NOP มีการประมาณค่าได้ใกล้เคียงกับค่าที่แท้จริงมากกว่าวิธี M สะท้อนให้เห็นว่าวิธี NOP

2.2) ปฏิสัมพันธ์ระหว่างวิธีการเชื่อมโยงคะแนนและระดับความสัมพันธ์ระหว่างมิติความสามารถ (METHOD X r) พบว่า วิธีการเชื่อมโยงคะแนนและโครงสร้างมิติความสามารถมีปฏิสัมพันธ์ต่อค่าดัชนี RMSE ในมิติที่ 2 ($F_{(2,228)}$= 4.108, p=.018) และค่าดัชนี BIAS ในมิติที่ 1 ($F_{(2,228)}$= 3.848, p=.023) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าวิธีการเชื่อมโยงคะแนน และโครงสร้างมิติความสามารรถ มีปฏิสัมพันธ์ที่ส่งผลต่อดัชนีบ่งชี้คุณภาพ RMSE ในมิติที่ 2 และดัชนี BIAS ในมิติที่ 1 นั่นเอง ซึ่งสามารถแสดงแผนภาพปฏิสัมพันธ์ระหว่าง 2 ตัวแปร ดังต่อไปนี้



แผนภาพที่ 4.6 กราฟแสดงปฏิสัมพันธ์ METHOD X r

จากแผนภาพ จะเห็นว่า เส้นกราฟของวิธี M และ วิธี NOP ไม่ขนานกัน จากภาพไม่ว่าจะเป็นโครงสร้างในลักษณะใดวิธี NOP จะมีค่า BIAS ที่ต่ำกว่าวิธี M แสดงว่าวิธี NOP มีการประมาณค่าได้ใกล้เคียงกับค่าที่แท้จริงมากกว่าวิธี M เมื่อพิจารณาค่า Ln RMSE พบว่าวิธี NOP จะมีค่าสูงกว่าวิธี M แสดงว่าวิธี NOP มีความคงที่ในการประมาณค่ามากกว่าวิธี M

2.3) ปฏิสัมพันธ์ระหว่างโครงสร้างมิติความสามารถและระดับความสัมพันธ์ระหว่างมิติความสามารถ (DS X r) พบว่า โครงสร้างมิติความสามารถและระดับความสัมพันธ์ระหว่างมิติความสามารถมีปฏิสัมพันธ์ที่ส่งผลต่อค่าดัชนีบ่งชี้คุณภาพอย่างมีนัยสำคัญทางสถิติเกือบทุกตัวแปร มีเพียงดัชนี BIAS ในมิติที่ 2 เท่านั้นที่ไม่มีนัยสำคัญทางสถิติ แสดงให้เห็นว่าโครงสร้างมิติความสามารถและระดับความสัมพันธ์ระหว่างมิติความสามารถมีปฏิสัมพันธ์ซึ่งส่งผลต่อดัชนี RMSE BIAS และ CORR แต่ไม่ส่งผลต่อดัชนี BIAS ในมิติที่ 2

3) ผลของอิทธิพลหลัก

3.1) เมื่อพิจารณาวิธีการเชื่อมโยงคะแนน (METHOD) พบว่า มีอิทธิพลต่อค่าดัชนี Ln RMSE และ BIAS ทั้งในมิติที่ 1 และมิติที่ 2 อย่างมีนัยสำคัญทางสถิติ แต่ไม่มีอิทธิพลต่อค่าCORR โดยเมื่อทำการเปรียบเทียบความแตกต่างของดัชนีบ่งชี้คุณภาพ ทั้งในมิติที่ 1 และ 2 พบว่า วิธี NOP มีความคงที่ในการประมาณค่ามากกว่าวิธี M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

ซึ่งพิจารณาจากค่า Ln RMSE และการประมาณค่าใกล้เคียงกับค่าที่แท้จริงมากกว่าวิธี M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ซึ่งพิจารณาจากดัชนี BIAS แต่เมื่อพิจารณาถึงค่า CORR พบว่าทั้ง 2 วิธี ไม่แตกต่างกัน

3.2) สำหรับโครงสร้างมิติความสามารถ (DS) พบว่ามีอิทธิพลต่อค่าดัชนี Ln RMSE ในมิติที่ 2 BIAS ในมิติที่ 1 และ CORR ทั้งในมิติที่ 1 และมิติที่ 2 โดยเมื่อทำการเปรียบเทียบความแตกต่างของดัชนีบ่งชี้คุณภาพ ทั้งในมิติที่ 1 และ 2 พบว่า โครงสร้างที่ไม่ซับซ้อน (APSS) มีความคงที่ในการประมาณค่าและการประมาณค่าใกล้เคียงกับค่าที่แท้จริงมากกว่าโครงสร้างที่ซับซ้อนมากขึ้นอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ซึ่งพิจารณาจากดัชนี Ln RMSE ในมิติที่ 2 และ BIAS ในมิติที่ 1 ตามลำดับ และเมื่อพิจารณาถึงค่า CORR พบว่าโครงสร้างที่ไม่ซับซ้อน (APSS) มีค่า CORR สูงกว่าโครงสร้างที่ซับซ้อนมากขึ้น (MS) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

3.3) ส่วนระดับความสัมพันธ์ระหว่างมิติความสามารถ (r) พบว่า มีอิทธิพลต่อค่าดัชนีบ่งชี้คุณภาพทั้งในมิติที่ 1 และมิติที่ 2 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยเมื่อทำการเปรียบเทียบความแตกต่างของดัชนีบ่งชี้คุณภาพ ทั้งในมิติที่ 1 และ 2 พบว่า เมื่อความสัมพันธ์เท่ากับ 0.5 จะมีความคงที่ในการประมาณค่าที่ดีกว่าระดับความสัมพันธ์ 0.0 และ 0.8 ตามลำดับ เมื่อพิจารณาถึงความถูกต้องในการประมาณค่าซึ่งพิจารณาจากดัชนี BIAS พบว่า เมื่อมีระดับความสัมพันธ์เท่ากับ 0.8 จะมีการประมาณค่าใกล้เคียงกับค่าที่แท้จริงมากกว่าระดับความสัมพันธ์ 0.5 และ 0.0 ตามลำดับ และเมื่อพิจารณาถึงค่า CORR พบว่าเมื่อความสัมพันธ์มีระดับสูงขึ้นจะมีค่า CORR ลดลงอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

โดยสรุป พบว่าวิธีการเชื่อมโยงคะแนน โครงสร้างมิติความสามารถ และระดับความสัมพันธ์ระหว่างมิติความสามารถ ไม่มีปฏิสัมพันธ์ที่ส่งผลต่อดัชนีบ่งชี้คุณภาพ RMSE และ BIAS แต่มีปฏิสัมพันธ์กันที่ส่งผลต่อค่า CORR นั่นคือ ในกรณีที่แบบสอบมีโครงสร้างที่ไม่ซับซ้อน (APSS) ผลของวิธีการเชื่อมโยงคะแนนและระดับความสัมพันธ์ระหว่างมิติความสามารถมีปฏิสัมพันธ์ร่วมกันในการส่งผลต่อค่า CORR ของค่าอำนาจจำแนกทั้ง 2 มิติ โดยเมื่อความสัมพันธ์ระหว่างมิติความสามารถเท่ากับ 0 และ 0.5 พบว่าวิธี M และวิธี NOP มีค่าCORR ใกล้เคียงกัน แสดงว่าความสัมพันธ์ระหว่างคะแนนที่ได้จากการแปลงคะแนนกับคะแนนใกล้เคียงกัน แต่เมื่อระดับความสัมพันธ์หว่างมิติความสามารถเท่ากับ 0.8 ในกรณีที่โครงสร้างมีความซับซ้อนมากขึ้น (MS) พบว่า วิธี NOP จะมีค่า CORR มากกว่าวิธี M แสดงว่าความสัมพันธ์ระหว่างค่าที่ได้จากการแปลงคะแนนกับคะแนนเกณฑ์ที่ได้จากวิธี NOP มีค่าสูงกว่าวิธี M

เมื่อเปรียบเทียบระหว่างวิธีการเชื่อมโยงคะแนน พบว่าวิธี NOP มีความคงที่ในการประมาณค่าและการประมาณค่าใกล้เคียงกับค่าที่แท้จริงมากกว่าวิธี M อย่างมีนัยสำคัญ

ทางสถิติที่ระดับ .05 แต่ความสัมพันธ์ระหว่างค่าที่ได้จากการแปลงคะแนนกับคะแนนเกณฑ์ทั้งสองวิธีแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

เมื่อเปรียบเทียบระหว่างโครงสร้างมิติความสามารถระหว่างโครงสร้างที่ไม่ซับซ้อน (APSS) และโครงสร้างที่ซับซ้อนมากขึ้น (MS) พบว่า โครงสร้างที่ไม่ซับซ้อน (APSS) มีความคงที่ในการประมาณค่าและการประมาณค่าใกล้เคียงกับค่าที่แท้จริงมากกว่าโครงสร้างที่ซับซ้อนมากขึ้น (MS) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยโครงสร้างที่ไม่ซับซ้อน (APSS) มีความสัมพันธ์ระหว่างค่าที่ได้จากจากการแปลงคะแนนกับคะแนนเกณฑ์ (CORR) สูงกว่าโครงสร้างที่ซับซ้อนมากขึ้น (MS) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

เมื่อเปรียบเทียบระหว่างระดับความสัมพันธ์ระหว่างมิติความสามารถ (r) พบว่า เมื่อความสัมพันธ์เท่ากับ 0.5 จะมีความคงที่ในการประมาณค่าที่ดีกว่าระดับความสัมพันธ์ 0.0 และ 0.8 ตามลำดับ เมื่อพิจารณาถึงความถูกต้องในการประมาณค่า พบว่า เมื่อมีระดับความสัมพันธ์เท่ากับ 0.8 จะมีการประมาณค่าใกล้เคียงกับค่าที่แท้จริงมากกว่าระดับความสัมพันธ์ 0.5 และ 0.0 ตามลำดับ และเมื่อพิจารณาถึงค่า CORR ที่สะท้อนถึงความสัมพันธ์ระหว่างค่าที่ได้จากจากการแปลงคะแนนกับคะแนนเกณฑ์ พบว่าเมื่อความสัมพันธ์มีระดับสูงขึ้นจะมีค่า CORR ลดลงอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

1.2 ผลการพิจารณาคุณภาพของการเชื่อมโยงคะแนนในภาพรวมของการแปลงค่าพารามิเตอร์ข้อสอบจากระดับที่ 2 ไปยังระดับที่ 3 (L2_3)

ดัชนีบ่งชี้คุณภาพของวิธีการเชื่อมโยงคะแนน ประกอบด้วย RMSE BIAS และ CORR ดังนั้นจึงจำเป็นต้องพิจารณาค่าสถิติเบื้องต้นของดัชนีดังกล่าวเพื่อศึกษาลักษณะของข้อมูลของผลที่ได้จากการแปลงค่าพารามิเตอร์ของข้อสอบจากระดับที่ 2 ไปยังระดับที่ 3 (L2_3) ก่อนทำการวิเคราะห์ความแปรปรวนแบบพหุเพื่อศึกษาปฏิสัมพันธ์ระหว่าง 3 ตัวแปร ต่อไป

1.2.1 ค่าสถิติเบื้องต้นของดัชนีบ่งชี้คุณภาพการเชื่อมโยงคะแนน

การวิเคราะห์ข้อมูลในเบื้องต้นนี้เป็นการวิเคราะห์ด้วยค่าสถิติพื้นฐานเพื่อบรรยายลักษณะของดัชนี BIAS RMSE และ CORR โดยการเสนอผลครั้งนี้มุ่งเน้นเปรียบเทียบผลที่ได้จากการแปลงค่าพารามิเตอร์อำนาจจำแนกในมิติที่ 1 (a_1) และค่าอำนาจจำแนกในมิติที่ 2 (a_2) เท่านั้น สำหรับผลที่ได้จากการแปลงค่าพารามิเตอร์ความยากของข้อสอบ (d) จะนำเสนอค่าดัชนีบ่งชี้คุณภาพเพื่อใช้สำหรับการพิจารณาระดับความมากน้อยของความคลาดเคลื่อนและความสัมพันธ์ที่เกิดขึ้นเพื่อเป็นสารสนเทศสำคัญในการนำไปเลือกใช้ในสถานการณ์ต่างๆ ได้อย่างเหมาะสม โดยมีรายละเอียดดังแสดงในตารางที่ 4.6

ตารางที่ 4.6 ค่าสถิติเบื้องต้นสำหรับการพิจารณาดัชนี RMSE BIAS และ CORR ของผลที่ได้จากการแปลงค่าพารามิเตอร์ข้อสอบจากระดับที่ 2 ไปยังระดับที่ 3

สถิติ	เงื่อนไขการจำลองข้อมูล			RMSE			BIAS			CORR		
				a1	a2	d	a1	a2	d	a1	a2	d
MIN	r = 0.0	APSS	M	0.086	0.098	0.189	-0.199	-0.201	0.695	0.888	0.903	0.913
			NOP	0.086	0.095	0.189	-0.181	-0.191	0.695	0.889	0.908	0.913
		MS	M	0.080	0.090	0.166	-0.297	-0.227	0.998	0.754	0.755	0.924
			NOP	0.079	0.089	0.166	-0.280	-0.248	0.998	0.728	0.826	0.924
	r = 0.5	APSS	M	0.054	0.060	0.165	-0.118	-0.122	0.937	0.676	0.767	0.732
			NOP	0.052	0.060	0.165	-0.115	-0.110	0.937	0.675	0.745	0.732
		MS	M	0.108	0.094	0.114	-0.154	-0.151	0.427	0.690	0.662	0.742
			NOP	0.107	0.086	0.114	-0.121	-0.109	0.427	0.692	0.637	0.742
	r = 0.8	APSS	M	0.097	0.102	0.143	-0.108	-0.108	0.990	0.530	0.567	0.955
			NOP	0.095	0.100	0.143	-0.091	-0.103	0.990	0.517	0.612	0.955
		MS	M	0.042	0.044	0.099	-0.192	-0.190	0.684	0.269	0.416	0.891
			NOP	0.042	0.043	0.099	-0.166	-0.170	0.684	0.365	0.419	0.891
MAX	r = 0.0	APSS	M	0.192	0.157	0.324	-0.100	-0.143	0.880	0.965	0.965	0.968
			NOP	0.191	0.156	0.324	-0.106	-0.153	0.880	0.965	0.965	0.968
		MS	M	0.207	0.167	0.284	-0.176	-0.190	1.059	0.955	0.934	0.977
			NOP	0.205	0.151	0.284	-0.180	-0.190	1.059	0.955	0.935	0.977
	r = 0.5	APSS	M	0.207	0.182	0.430	-0.078	-0.078	1.006	0.982	0.972	0.978
			NOP	0.207	0.179	0.430	-0.072	-0.080	1.006	0.981	0.973	0.978
		MS	M	0.199	0.156	0.218	-0.082	-0.070	0.495	0.980	0.983	0.986
			NOP	0.172	0.143	0.218	0.002	-0.017	0.495	0.986	0.983	0.986
	r = 0.8	APSS	M	0.207	0.160	0.228	-0.042	-0.032	1.148	0.839	0.848	0.981
			NOP	0.177	0.153	0.228	-0.038	-0.028	1.148	0.838	0.850	0.981
		MS	M	0.217	0.225	0.303	-0.032	-0.029	1.134	0.970	0.979	0.991
			NOP	0.198	0.186	0.303	-0.031	-0.024	1.134	0.971	0.980	0.991
MEAN	r = 0.0	APSS	M	0.132	0.124	0.241	-0.162	-0.172	0.821	0.933	0.932	0.946
			NOP	0.130	0.123	0.241	-0.156	-0.174	0.821	0.933	0.933	0.946
		MS	M	0.120	0.121	0.209	-0.211	-0.207	1.029	0.899	0.888	0.962
			NOP	0.118	0.119	0.209	-0.208	-0.207	1.029	0.897	0.891	0.962
	r = 0.5	APSS	M	0.110	0.119	0.222	-0.098	-0.098	0.963	0.904	0.887	0.948
			NOP	0.108	0.117	0.222	-0.091	-0.094	0.963	0.901	0.884	0.948
		MS	M	0.134	0.130	0.181	-0.115	-0.115	0.461	0.827	0.787	0.957
			NOP	0.129	0.122	0.181	-0.062	-0.073	0.461	0.834	0.794	0.957
	r = 0.8	APSS	M	0.141	0.139	0.183	-0.067	-0.072	1.030	0.739	0.733	0.973
			NOP	0.132	0.131	0.183	-0.055	-0.060	1.030	0.727	0.731	0.973
		MS	M	0.167	0.156	0.159	-0.130	-0.094	0.737	0.570	0.648	0.974
			NOP	0.142	0.137	0.159	-0.110	-0.077	0.737	0.613	0.665	0.974
SD	r = 0.0	APSS	M	0.025	0.016	0.041	0.025	0.015	0.046	0.019	0.016	0.018
			NOP	0.025	0.016	0.041	0.019	0.011	0.046	0.019	0.017	0.018
		MS	M	0.028	0.018	0.038	0.025	0.011	0.019	0.048	0.039	0.015
			NOP	0.027	0.016	0.038	0.021	0.014	0.019	0.053	0.027	0.015
	r = 0.5	APSS	M	0.031	0.026	0.057	0.013	0.013	0.017	0.062	0.045	0.056
			NOP	0.031	0.026	0.057	0.012	0.008	0.017	0.064	0.050	0.056
		MS	M	0.020	0.017	0.023	0.021	0.018	0.019	0.063	0.073	0.052
			NOP	0.015	0.018	0.023	0.043	0.028	0.019	0.073	0.074	0.052
	r = 0.8	APSS	M	0.026	0.018	0.024	0.019	0.016	0.032	0.086	0.075	0.008
			NOP	0.022	0.017	0.024	0.014	0.015	0.032	0.089	0.070	0.008
		MS	M	0.041	0.037	0.043	0.041	0.045	0.096	0.166	0.129	0.021
			NOP	0.031	0.030	0.043	0.038	0.041	0.096	0.155	0.131	0.021

จากตารางที่ 4.6 *เมื่อพิจารณาเงื่อนไขไม่มีความสัมพันธ์ระหว่างมิติสำหรับโครงสร้างมิติความสามารถที่ไม่ซับซ้อน (APSS_0.0)* ของผลที่ได้จากการแปลงค่าพารามิเตอร์ข้อสอบจากระดับที่ 2 ไประดับที่ 3 พบว่า ค่าต่ำสุดของดัชนี RMSE (RMSE_a_1=0.086, RMSE_a_2=0.098) เท่ากันกับวิธี NOP (RMSE_a_1=0.086, RMSE_a_2=0.095) สำหรับค่าพารามิเตอร์อำนาจจำแนกในมิติที่ 1 และมีค่าสูงกว่าเพียงเล็กน้อยสำหรับค่าพารามิเตอร์อำนาจจำแนกในมิติที่ 2 แสดงว่าทั้ง 2 วิธี มีความคลาดเคลื่อนของการประมาณค่าของค่าต่ำสุดใกล้เคียงกัน สำหรับค่าสูงสุด พบว่า วิธี M (RMSE_a_1=0.192, RMSE_a_2=0.157) มีความคลาดเคลื่อนของการประมาณค่าใกล้เคียงกับวิธี NOP (RMSE_a_1=0.191, RMSE_a_2=0.156) ในทุกค่าพารามิเตอร์ เมื่อพิจารณาค่า BIAS พบว่า วิธี M ให้ค่าต่ำสุดของ BIAS (BIAS_a_1=-0.199, BIAS_a_2=-0.201) ต่ำกว่าความเป็นจริงมากกว่าวิธี NOP (BIAS_a_1= -0.181, BIAS_a_2= -0.191) ในขณะที่วิธี M ให้ค่าสูงสุด (BIAS_a_1= -0.100, BIAS_a_2=-0.143) ใกล้เคียงกับค่าที่แท้จริงมากกว่าวิธี NOP (BIAS_a_1=-0.106, BIAS_a_2= -0.153) สำหรับการพิจารณาค่าเฉลี่ยของ RMSE และ BIAS พบว่าวิธี M (RMSE_a_1=0.132, RMSE_a_2=0.124, BIAS_a_1= -0.162, BIAS_a_2= -0.172) มีความคลาดเคลื่อน และมีการประมาณค่าต่ำกว่าความเป็นจริงมากกว่าวิธี NOP (RMSE_a_1=0.130, RMSE_a_2=0.123, BIAS_a_1= -0.156, BIAS_a_2= -0.174) อย่างไรก็ตามพบว่าวิธี NOP มีการประมาณค่าต่ำกว่าความเป็นจริงมากกว่าวิธี M สำหรับค่าอำนาจจำแนกในมิติที่ 2 (a_2) เมื่อพิจารณาค่า CORR พบว่าทั้งสองวิธีมีค่าไม่แตกต่างกันและเข้าใกล้ 1 แสดงว่าค่าพารามิเตอร์ที่ได้จากการแปลงคะแนนโดยวิธี M และวิธี NOP มีความสัมพันธ์กับคะแนนเกณฑ์สูงใกล้เคียงกัน นั่นเอง

เมื่อพิจารณาเงื่อนไขไม่มีความสัมพันธ์ระหว่างมิติ สำหรับโครงสร้างมิติที่ซับซ้อนมากขึ้น (MS_0.0) พบว่า วิธี M ให้ค่าต่ำสุด (RMSE_a_1=0.080, RMSE_a_2=0.090) และสูงสุด (RMSE_a_1=0.207, RMSE_a_2=0.167) ของ RMSE มากกว่าวิธี NOP แสดงว่าเกิดความคลาดเคลื่อนของการประมาณค่ามากกว่า ในขณะที่วิธี M ให้ค่าต่ำสุดที่ต่ำกว่าความเป็นจริง (BIAS_a_1= -0.297, BIAS_a_2= -0.227) มากกว่าวิธี NOP ใน a_1 แต่ประมาณค่าได้ใกล้เคียงกับค่าที่แท้จริงใน a_2 (BIAS_a_1= -0.280, BIAS_a_2= -0.248) เมื่อพิจารณาค่าเฉลี่ยพบว่า วิธี M (RMSE_a_1=0.207, RMSE_a_2=0.167) มีความคลาดเคลื่อนของการประมาณค่ามากกว่าวิธี NOP (RMSE_a_1=0.205, RMSE_a_2=0.151) ซึ่งวิธี M ให้ค่าเฉลี่ยของ BIAS (BIAS_a_1= -0.176, BIAS_a_2= -0.190) ที่ต่ำกว่าความเป็นจริงมากกว่าวิธี NOP (BIAS_a_1= -0.180, BIAS_a_2= -0.190) ใน a_1 แต่มีค่าที่ไม่แตกต่างกันใน a_2 เมื่อพิจารณาค่าสูงสุดของค่าสัมประสิทธิ์สหสัมพันธ์ พบว่าวิธี M มีค่ามากกว่าวิธี NOP ทุกค่าพารามิเตอร์ สำหรับค่าสัมประสิทธิ์สหสัมพันธ์โดยเฉลี่ยของทั้ง 2 วิธี พบว่ามีค่าสูงและใกล้เคียงกันเหมือนเงื่อนไข MS_0.0

เมื่อพิจารณาเงื่อนไขความสัมพันธ์ระหว่างมิติในระดับปานกลางสำหรับโครงสร้างที่ไม่ซับซ้อน (APSS_0.5) พบว่า ค่าต่ำสุดของ RMSE ไม่แตกต่างกันทั้ง 2 วิธี สำหรับค่า BIAS พบว่า วิธี M (BIAS_a_1= -0.118, BIAS_a_2= -0.122) ให้ค่าสูงกว่า NOP (BIAS_a_1= -0.115, BIAS_a_2= -0.110) แสดงว่าวิธี M มีประมาณค่าสูงกว่าความเป็นจริงมากกว่าวิธี NOP เมื่อพิจารณาค่าสูงสุดของดัชนี RMSE พบว่า วิธี M (RMSE_a_1=0.207, RMSE_a_2=0.182) และ วิธี NOP (RMSE_a_1=0.207, RMSE_a_2=0.179) มีค่าไม่แตกต่างกัน ใน a1 แต่วิธี M มีค่า RMSE มากกว่าวิธี NOP ใน a_2 สำหรับค่า BIAS พบว่า วิธี M (BIAS_a_1= -0.078, BIAS_a_2= -0.078) ประมาณค่าได้สูงกว่าความเป็นจริงมากกว่าวิธี NOP (BIAS_a_1= -0.072, BIAS_a_2= -0.080) ใน a1 แต่ประมาณค่าได้ใกล้เคียงกับค่าที่แท้จริงมากกว่า NOP ใน a_2 สำหรับค่า CORR ของทั้ง 2 วิธี พบว่ามีค่าสัมประสิทธิ์สหสัมพันธ์อยู่ในระดับสูงเหมือนเงื่อนไข APSS_0.0 และ MS_0.0

เมื่อพิจารณาเงื่อนไขความสัมพันธ์ระหว่างมิติในระดับปานกลางสำหรับโครงสร้างที่ซับซ้อนมากขึ้น (MS_0.5) พบว่า วิธี M มีค่าต่ำสุด (RMSE_a_1=0.108, RMSE_a_2=0.094, BIAS_a_1= -0.154, BIAS_a_2= -0.151) และสูงสุด (RMSE_a_1=0.199, RMSE_a_2=0.156, BIAS_a_1=-0.082, BIAS_a_2= -0.070) ของดัชนี RMSE และ BIAS สูงกว่า วิธี NOP ในทุกค่าพารามิเตอร์ แสดงว่าวิธี NOP มีความคงที่ของการประมาณค่าและมีการประมาณค่าใกล้เคียงกับค่าที่แท้จริงมากกว่าวิธี M ซึ่งให้ผลเช่นเดียวกับการพิจารณาค่าเฉลี่ยของดัชนี RMSE และ BIAS ที่พบว่า วิธี M (RMSE_a_1=0.134, RMSE_a_2=0.130, BIAS_a_1= -0.115, BIAS_a_2= -0.115) มีความคลาดเคลื่อนสูงและมีการประมาณค่าต่ำกว่าความเป็นจริงมากกว่าวิธี NOP (RMSE_a_1=0.129, RMSE_a_2=0.122, BIAS_a_1= -0.062, BIAS_a_2= -0.073) เมื่อพิจารณาค่า CORR ของทั้ง 2 วิธี พบว่าอยู่ในช่วงปานกลางค่อนข้างสูง

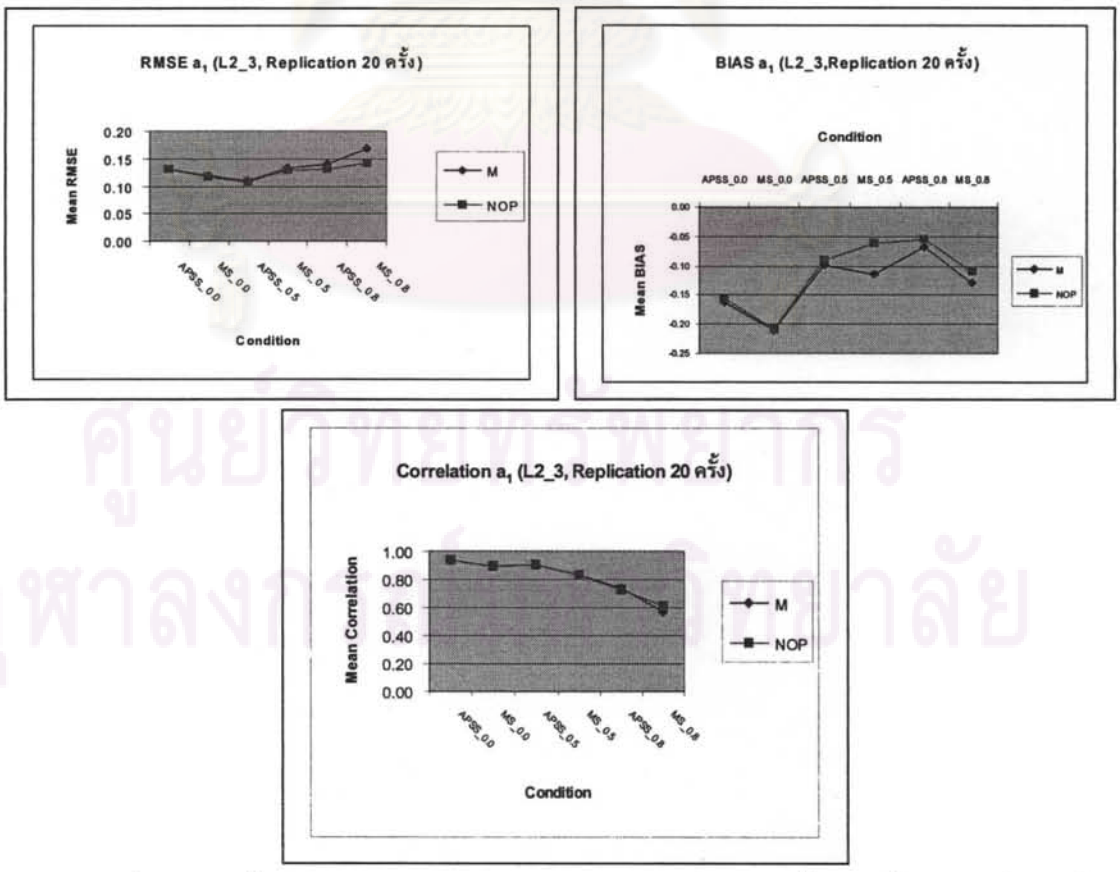
เมื่อพิจารณาเงื่อนไขความสัมพันธ์ระหว่างมิติในระดับสูง สำหรับโครงสร้างที่ไม่ซับซ้อน (APSS_0.8) พบว่า มีลักษณะใกล้เคียงกับเงื่อนไข MS_0.5 แต่ต่างกันในประเด็นของค่า CORR ของทั้ง 2 วิธี อยู่ในระดับปานกลางไม่สูงเหมือนเงื่อนไขที่มีความสัมพันธ์ระหว่างมิติในระดับปานกลางหรือไม่มีความสัมพันธ์ระหว่างมิติ แสดงว่าเมื่อความสัมพันธ์ระหว่างมิติความสามารถสูงขึ้นส่งผลทำให้ค่าพารามิเตอร์ที่ได้จากการแปลงคะแนนมีความสัมพันธ์กับคะแนนเกณฑ์ต่ำลง

เมื่อพิจารณาเงื่อนไขความสัมพันธ์ระหว่างมิติในระดับสูง สำหรับโครงสร้างที่ซับซ้อนมากขึ้น (MS_0.8) พบว่า ทั้งสองวิธีมีค่าต่ำสุดของ RMSE ใกล้เคียงกันมาก แต่ค่า BIAS ที่ต่ำสุดของวิธี M (BIAS_a_1 = -0.192, BIAS_a_2 = -0.190) มีค่าสูงกว่าวิธี NOP (BIAS_a_1= -0.166, BIAS_a_2= -0.170) นั่นคือวิธี M มีการประมาณค่าต่ำกว่าความเป็นจริงมากกว่าวิธี NOP สำหรับค่าสูงสุดและค่าเฉลี่ยของ RMSE และ BIAS พบว่าวิธี M มีค่าสูงกว่าวิธี NOP โดยมีการประมาณ

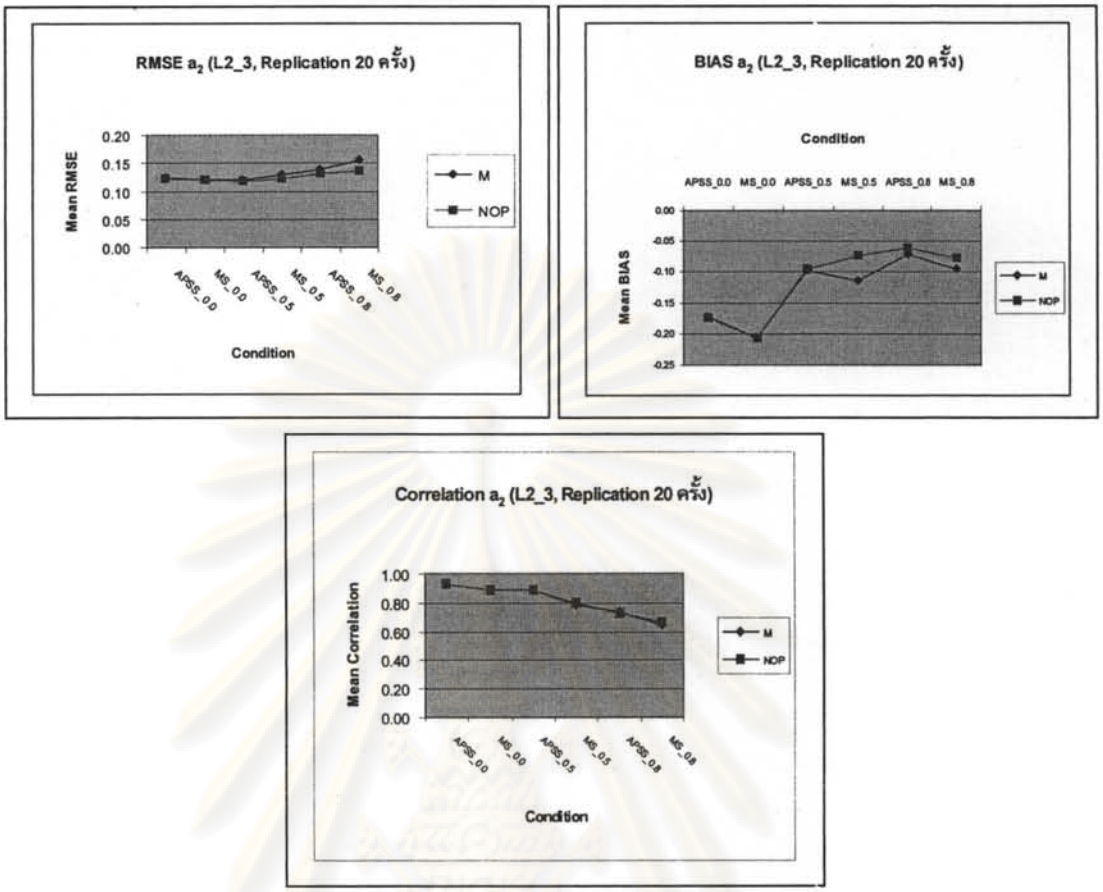
ค่าที่สูงกว่าความเป็นจริง เมื่อพิจารณาค่า CORR พบว่าอยู่ในช่วงระดับปานกลาง (0.4-0.6) โดยวิธี NOP มีค่าสูงกว่าวิธี M แสดงว่าเมื่อความสัมพันธ์ระหว่างมิติความสามารถสูงขึ้นส่งผลทำให้ค่าพารามิเตอร์ที่ได้จากการแปลงคะแนนโดยวิธี NOP มีความสัมพันธ์กับคะแนนเกณฑ์สูงกว่าวิธี M แต่มีค่าในระดับไม่สูงเหมือนเงื่อนไข APSS_0.0, MS_0.0, APSS_0.5, MS_0.5 และ MS_0.8

เมื่อพิจารณาค่าพารามิเตอร์ d โดยภาพรวม พบว่า มีลักษณะใกล้เคียงกับผลที่ได้จากการแปลงค่าพารามิเตอร์ของข้อสอบระดับที่ 1 ไป 2 แต่แตกต่างกันในประเด็นของค่า BIAS ที่พบว่าเมื่อโครงสร้างมิติความสามารถมีความซับซ้อนมากขึ้นส่งผลทำให้ประมาณค่ามากกว่าความเป็นจริงเพิ่มขึ้น แต่อย่างไรก็ตามผลการวิจัยพบว่า ทั้งสองวิธีประมาณค่าได้ใกล้เคียงกับค่าที่แท้จริงเมื่อความสัมพันธ์ของมิติความสามารถอยู่ในระดับปานกลาง (Corr 0.5) และมีโครงสร้างมิติความสามารถที่ซับซ้อนมากขึ้น (MS) สำหรับค่า CORR พบว่ามีค่าใกล้เคียงกัน คือมีค่าเข้าใกล้ 1 อย่างไรก็ตามพบว่าเมื่อความสัมพันธ์ระหว่างมิติความสามารถอยู่ในระดับสูงมีผลทำให้ค่า CORR ลดลง

จากที่กล่าวมาเป็นการบรรยายถึงค่าสถิติเบื้องต้นของดัชนี RMSE BIAS และ CORR ทั้ง 6 เงื่อนไข คือ APSS_0.0, MS_0.0, APSS_0.0, MS_0.5, APSS_0.8 และ MS_0.8 โดยมุ่งเน้นการเปรียบเทียบระหว่างวิธี M และ NOP เพื่อให้ง่ายต่อความเข้าใจสามารถสรุปรวมทั้ง 6 เงื่อนไข ในแต่ละค่าพารามิเตอร์ของการแปลงคะแนน ดังแผนภาพต่อไปนี้



แผนภาพที่ 4.7 ค่าเฉลี่ยดัชนี RMSE BIAS และ CORR ของ a_1 ในการทำซ้ำ 20 ครั้ง ของ 6 เงื่อนไขที่ศึกษาเปรียบเทียบระหว่างวิธี M และ NOP

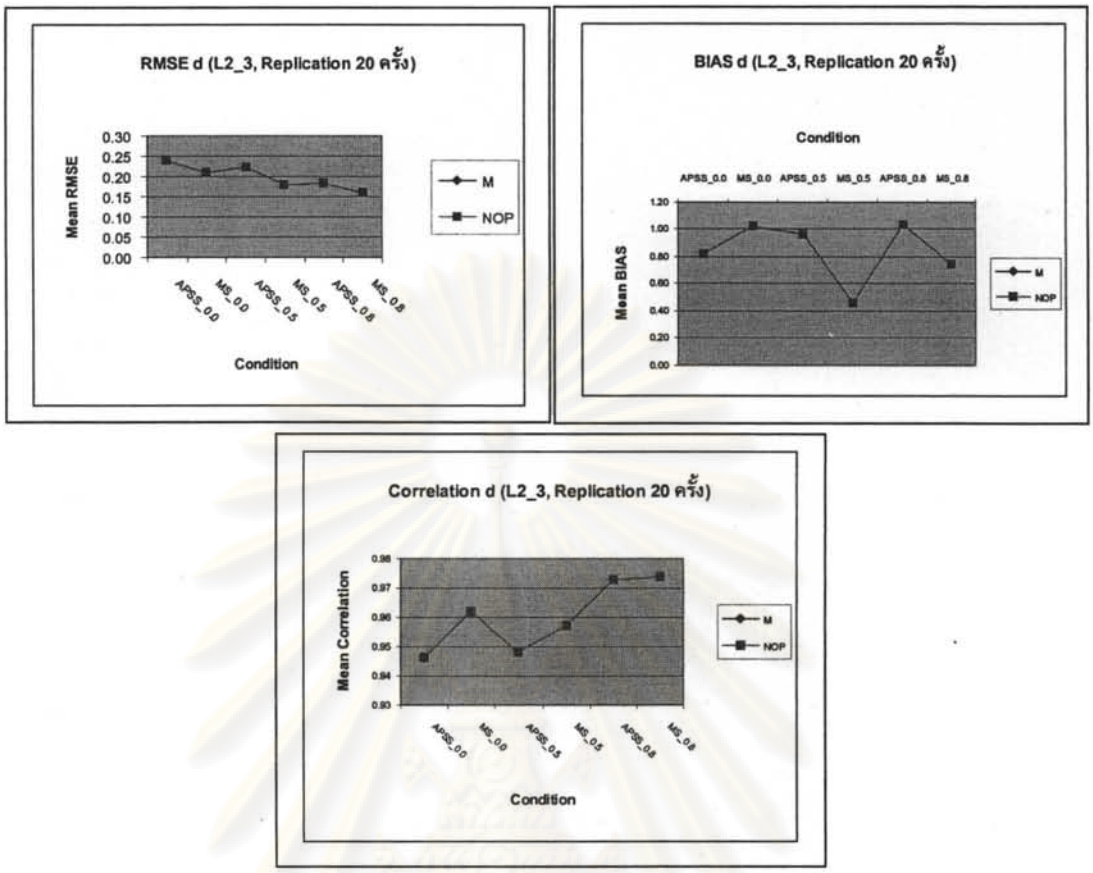


แผนภาพที่ 4.8 ค่าเฉลี่ยดัชนี RMSE BIAS และ CORR ของ a_2 ในการทำซ้ำ 20 ครั้ง ของ 6 เงื่อนไข ที่ศึกษาเปรียบเทียบระหว่างวิธี M และ NOP

จากแผนภาพที่ 4.7 และ 4.8 ให้ผลการแปลงคะแนนของค่าพารามิเตอร์อำนาจจำแนกในมิติที่ 1 และ 2 ใกล้เคียงกัน ซึ่งมีลักษณะเหมือนการแปลงคะแนนจากระดับที่ 1 ไปยังระดับที่ 2 (L1_2) นั่นคือ เมื่อพิจารณาจากดัชนี RMSE พบว่า การแปลงค่าพารามิเตอร์ในมิติที่ 1 มีความคงที่ของการแปลงคะแนนระหว่างวิธี M และวิธี NOP ใกล้เคียงกันเมื่อศึกษากับเงื่อนไขที่ไม่มีระดับความสัมพันธ์หรือความสัมพันธ์ในระดับกลาง แต่เมื่อความสัมพันธ์ระหว่างมิติความสามารถสูงขึ้นวิธี NOP จะมีความคงที่มากกว่าวิธี M อย่างเห็นได้ชัด

เมื่อพิจารณาจากดัชนี BIAS พบว่าทั้งสองวิธี มีการประมาณค่าที่ต่ำกว่าความเป็นจริง เนื่องจากมีค่าดัชนีติดลบ อย่างไรก็ตามวิธี NOP มีการประมาณค่าที่ใกล้เคียงกับค่าที่แท้จริงมากกว่าวิธี M อย่างเห็นได้ชัด โดยโครงสร้างที่ไม่ซับซ้อน (APSS) จะมีการประมาณค่าใกล้เคียงกับค่าที่แท้จริงมากกว่าโครงสร้างที่ซับซ้อนมากขึ้น (MS)

เมื่อพิจารณาจากค่า CORR พบว่า ทั้งสองวิธีมีค่าใกล้เคียงกันมากและมีค่าเข้าใกล้ 1 แต่เมื่อความสัมพันธ์ระหว่างมิติความสามารถสูง พบว่า วิธี NOP มีค่าสูงกว่าวิธี M แสดงว่าวิธี NOP มีค่าความสัมพันธ์ของคะแนนที่แปลงได้กับคะแนนเกณฑ์สูงกว่าวิธี M



แผนภาพที่ 4.9 ค่าเฉลี่ยดัชนี RMSE BIAS และ CORR ของ d ในการทำซ้ำ 20 ครั้ง ของ 6 เงื่อนไขที่ศึกษา เปรียบเทียบระหว่างวิธี M และ NOP

จากแผนภาพที่ 4.9 เมื่อพิจารณาการแปลงค่าพารามิเตอร์ความยาก (d) พบว่า มีค่าดัชนี RMSE BIAS และ CORR เท่ากันทั้งวิธี M และวิธี NOP โดยเมื่อพิจารณา RMSE พบว่า เมื่อความสัมพันธ์ระหว่างมิติเพิ่มขึ้นและมีโครงสร้างที่ซับซ้อนมากขึ้น พบว่ามีความคงที่ของการแปลงคะแนนมากกว่าซึ่งเมื่อเปรียบเทียบระหว่างโครงสร้างที่ซับซ้อนกับไม่ซับซ้อน พบว่าโครงสร้างที่ซับซ้อนมีความคงที่ของการเชื่อมโยงคะแนนมากกว่าโครงสร้างที่ไม่ซับซ้อน โดยจากแผนภาพพบว่าเงื่อนไข MS_0.8 มีความคงที่ของการแปลงคะแนนมากกว่าเงื่อนไขอื่น

เมื่อพิจารณาดัชนี BIAS พบว่า เมื่อความสัมพันธ์ระหว่างมิติเพิ่มขึ้น มีการประมาณค่าที่ใกล้เคียงกับค่าที่แท้จริงมากกว่า ซึ่งเมื่อเปรียบเทียบระหว่างโครงสร้างที่ซับซ้อนกับไม่ซับซ้อน พบว่าโครงสร้างที่ไม่ซับซ้อนมีการประมาณค่าใกล้เคียงกับค่าที่แท้จริงมากกว่าโครงสร้างที่ซับซ้อน โดยจากแผนภาพพบว่าเงื่อนไข MS_0.5 มีความถูกต้องของการแปลงคะแนนมากกว่าเงื่อนไขอื่น

เมื่อพิจารณา CORR พบว่า เมื่อความสัมพันธ์ระหว่างมิติความสามารถเพิ่มขึ้น ค่าความสัมพันธ์ระหว่างค่าที่ได้จากการแปลงคะแนนกับคะแนนเกณฑ์มีค่าสูงขึ้น แต่อย่างไรก็ตาม พบว่าทั้ง 6 เงื่อนไขที่ศึกษามีค่าเข้าใกล้ 1

โดยสรุป เมื่อพิจารณาจากค่าสถิติเบื้องต้นพบว่าผลการแปลงคะแนนทั้งจากระดับที่ 1 ไปยังระดับที่ 2 และผลการแปลงคะแนนจากระดับที่ 2 ไปยังระดับที่ 3 ให้ผลใกล้เคียงกัน นั่นคือ วิธี M และ วิธี NOP มีค่าดัชนีบ่งชี้คุณภาพใกล้เคียงกันเมื่อศึกษาในเงื่อนไขไม่มีความสัมพันธ์หรือความสัมพันธ์ระหว่างมิติความสามารถปานกลาง (r=0.0, r=0.5) แต่เมื่อความสัมพันธ์ระหว่างมิติเพิ่มขึ้น (r=0.8) และโครงสร้างเชิงมิติมีความซับซ้อนมากขึ้น (MS) พบว่า วิธี NOP มีความคงที่และมีความถูกต้องในการประมาณค่ามากกว่าวิธี M โดยวิธี M แนวโน้มในการประมาณค่าที่ต่ำกว่าความเป็นจริง (Overestimate)

เมื่อพิจารณาถึงค่า CORR พบว่ามีค่าความสัมพันธ์สูงทั้งสองวิธี โดยมีค่าเข้าใกล้ 1 อย่างไรก็ตามเมื่อความสัมพันธ์ระหว่างมิติความสามารถเพิ่มขึ้นพบว่า NOP มีค่าความสัมพันธ์ของคะแนนที่แปลงได้กับคะแนนเกณฑ์สูงกว่าวิธี M

ต่อไปเป็นขั้นตอนสำคัญสำคัญในการทดสอบสมมติฐานทางสถิติเพื่อศึกษาปฏิสัมพันธ์ระหว่างวิธีการเชื่อมโยงคะแนน โครงสร้างมิติความสามารถ และระดับความสัมพันธ์ระหว่างมิติความสามารถ ที่ส่งผลต่อคุณภาพของการเชื่อมโยงคะแนน โดยการวิเคราะห์ความแปรปรวนแบบพหุ (MANOVA) ซึ่งมีรายละเอียดดังนี้

1.2.2 ผลการวิเคราะห์ความแปรปรวนแบบพหุจำแนก 3 ทาง (Three-way MANOVA)

การนำเสนอผลการวิเคราะห์ข้อมูลในส่วนนี้มีลักษณะคล้ายคลึงกับการแปลงคะแนนจากระดับที่ 1 ไปยังระดับที่ 2 เพื่อตรวจสอบความแตกต่างของเซ็นทรอยด์ของตัวแปรดัชนีบ่งชี้คุณภาพ RMSE BIAS และ ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าที่แปลงได้จากแบบสอบปรับเทียบคะแนนและค่าที่ประมาณได้จากแบบสอบฐานของค่าพารามิเตอร์อำนาจจำแนกในมิติที่ 1 และมิติที่ 2 ซึ่งเป็นตัวแปรตามรวมทั้งหมด 6 ตัวแปร คือ 1) ค่าดัชนี Ln RMSE ในมิติที่ 1 (Ln RMSE_a1) 2) ค่าดัชนี Ln RMSE ในมิติที่ 2 (Ln RMSE_a2) 3) ค่าดัชนี BIAS ในมิติที่ 1 (BIAS_a1) 4) ค่าดัชนี BIAS ในมิติที่ 2 (BIAS_a2) 5) ค่าสัมประสิทธิ์สหสัมพันธ์ในมิติที่ 1 (CORR $_a_1$) และ 6) ค่าสัมประสิทธิ์สหสัมพันธ์ในมิติที่ 2 (CORR_a_2) ได้ผลการวิเคราะห์ดังนี้

ตารางที่ 4.7 การทดสอบ Box's Test และ Bartlett's Test

Box's M	1316.587	Bartlett's Test of Sphericity	
F	4.979597	Likelihood Ratio	.000
df1	231	Approx. Chi-Square	2068.264
df2	42086.41	df	20
p	.000	p	.000

จากตารางที่ 4.7 ผลการตรวจสอบข้อตกลงเบื้องต้นของสถิติวิเคราะห์ พบว่า เมตริกซ์ความแปรปรวน-ความแปรปรวนร่วมของตัวแปรทั้ง 6 ตัว คือ Ln RMSE_a1 Ln RMSE_a2 BIAS_a1 BIAS_a2 CORR_a_1 และ CORR_a_2 ต่างกันระหว่างกลุ่มอย่างมีนัยสำคัญทางสถิติ (Box's M =1316.587,df=(231, 42086.41), p=.000) ซึ่งเป็นการละเมิดข้อตกลงเบื้องต้นทางสถิติของการวิเคราะห์ความแปรปรวนของตัวแปรพหุนาม ที่กำหนดให้เมตริกซ์ความแปรปรวน-ความแปรปรวนร่วมของทุกกลุ่มต้องเท่ากัน ซึ่งการทดสอบด้วยวิธี Box's M ค่อนข้างมีความไวต่อการละเมิดข้อตกลงเบื้องต้นดังกล่าว ดังที่ได้กล่าวมาแล้วข้างต้น อย่างไรก็ตามจากงานวิจัยของ Holloway & Dunn (1967) Hakstain, Roed & Linn (1979) และ Olson (1974) พบว่า การที่เมตริกซ์ค่าแปรปรวนร่วมไม่เท่ากันจะไม่มีผลกระทบต่อระดับนัยสำคัญในแต่ละกลุ่มย่อยเท่ากัน ซึ่งในการวิจัยครั้งนี้ได้กำหนดให้ขนาดตัวอย่างในแต่ละกลุ่มย่อยเท่ากัน ซึ่งเรียกว่า Balanced Three-way MANOVA (อ้างใน กัลยา วานิชย์บัญชา, 2550) ซึ่งการทดสอบความแปรปรวนพหุนามด้วย F-test มีความแกร่งเพียงพอ เมื่อมีการละเมิดข้อตกลงเบื้องต้นดังกล่าว (Huynh & Mandeville, 1979 อ้างใน ณัฏภรณ์ หลาวทอง และ สกล ซื่อธนาพรกุล, 2550)

อย่างไรก็ดี การตรวจสอบความสัมพันธ์ระหว่างดัชนีบ่งชี้คุณภาพทั้ง 6 ตัวแปร ว่ามีความสัมพันธ์แตกต่างจากเมทริกซ์เอกลักษณ์หรือไม่ ด้วยการทดสอบ Bartlett's test พบว่า ทั้ง 6 ตัวแปรมีความสัมพันธ์กัน (Likelihood=.000, Approx Chi-square=2068.264, df=20, p=.000) ด้วยเหตุนี้จึงสามารถวิเคราะห์ความแปรปรวนตัวแปรพหุนามได้

สถิติทดสอบที่ใช้ในการทดสอบความแปรปรวนตัวแปรพหุนาม 4 ชนิด สำหรับงานวิจัยครั้งนี้จะใช้สถิติทดสอบทั้ง 4 ชนิดเป็นหลักฐานในการทดสอบ อย่างไรก็ตามผู้วิจัยเน้นที่สถิติทดสอบ Wilks' Lamda เนื่องจากเหมาะสมสำหรับนำมาใช้ในงานวิจัยนี้มากที่สุด ซึ่งมีตัวแปรอิสระจำนวน 3 กลุ่ม ดังที่ได้กล่าวมาแล้วข้างต้น

ตารางที่ 4.8 การวิเคราะห์ความแปรปรวนพหุนามดัชนีบ่งชี้คุณภาพของการเชื่อมโยงคะแนนทั้ง 6 ตัวแปร ระหว่างวิธีการเชื่อมโยงคะแนน โครงสร้างมิติความสามารถและระดับความสัมพันธ์ระหว่างมิติความสามารถที่ต่างกัน

Effect		Value	F	Hypothesis df	Error df	p
Intercept	Pillai's Trace	0.997	12167.501	6	223	.000
	Wilks' Lambda	0.003	12167.501	6	223	.000
	Hotelling's Trace	327.377	12167.501	6	223	.000
	Roy's Largest Root	327.377	12167.501	6	223	.000
วิธีการเชื่อมโยงคะแนน (METHOD)	Pillai's Trace	0.215	10.184	6	223	.000
	Wilks' Lambda	0.785	10.184	6	223	.000
	Hotelling's Trace	0.274	10.184	6	223	.000
	Roy's Largest Root	0.274	10.184	6	223	.000
โครงสร้างมิติความสามารถ (DS)	Pillai's Trace	0.521	40.363	6	223	.000
	Wilks' Lambda	0.479	40.363	6	223	.000
	Hotelling's Trace	1.086	40.363	6	223	.000
	Roy's Largest Root	1.086	40.363	6	223	.000
ระดับความสัมพันธ์ระหว่างมิติความสามารถ (r)	Pillai's Trace	1.446	97.483	12	448	.000
	Wilks' Lambda	0.027	187.627	12	446	.000
	Hotelling's Trace	18.260	337.816	12	444	.000
	Roy's Largest Root	17.257	644.245	6	224	.000
METHOD X DS	Pillai's Trace	0.066	2.644	6	223	.017
	Wilks' Lambda	0.934	2.644	6	223	.017
	Hotelling's Trace	0.071	2.644	6	223	.017
	Roy's Largest Root	0.071	2.644	6	223	.017
METHOD X r	Pillai's Trace	0.127	2.532	12	448	.003
	Wilks' Lambda	0.875	2.574	12	446	.003
	Hotelling's Trace	0.141	2.615	12	444	.002
	Roy's Largest Root	0.126	4.705	6	224	.000
DS X r	Pillai's Trace	0.507	12.686	12	448	.000
	Wilks' Lambda	0.524	14.179	12	446	.000
	Hotelling's Trace	0.849	15.707	12	444	.000
	Roy's Largest Root	0.772	28.817	6	224	.000
METHOD X DS X r	Pillai's Trace	0.102	2.009	12	448	.022
	Wilks' Lambda	0.898	2.045	12	446	.019
	Hotelling's Trace	0.112	2.081	12	444	.017
	Roy's Largest Root	0.107	3.996	6	224	.001

การทดสอบความแปรปรวนพหุนามของดัชนี RMSE BIAS และ CORR ของค่าอำนาจจำแนกในมิติที่ 1 และ มิติที่ 2 ทั้ง 6 ตัวแปร ในวิธีการเชื่อมโยงคะแนน โครงสร้างมิติความสามารถและระดับความสัมพันธ์ระหว่างมิติความสามารถที่ต่างกัน ผลการวิเคราะห์ พบว่าให้ผลที่ต่างจากการแปลงค่าพารามิเตอร์จากระดับที่ 1 ไปยังระดับที่ 2 นั่นคือ มีปฏิสัมพันธ์ระหว่าง 3 ตัวแปร คือ วิธีการเชื่อมโยงคะแนน โครงสร้างมิติความสามารถและระดับความสัมพันธ์ระหว่างมิติความสามารถ ซึ่งส่งผลทำให้มีค่าดัชนี Ln RMSE BIAS และค่า CORR ของค่าอำนาจจำแนกในมิติที่ 1 และ มิติที่ 2 โดยเฉลี่ยต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าวิธีการเชื่อมโยงคะแนน โครงสร้างมิติความสามารถและระดับความสัมพันธ์ระหว่างมิติความสามารถ ร่วมกันส่งผลทำให้มีดัชนีบ่งชี้คุณภาพที่ต่างกัน

ส่วนปฏิสัมพันธ์ระหว่าง 2 ตัวแปร พบว่ามีปฏิสัมพันธ์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ในทุกประเภทของสถิติทดสอบที่ศึกษา แสดงว่าปฏิสัมพันธ์ระหว่างวิธีการเชื่อมโยงคะแนนและระดับความสัมพันธ์ระหว่างมิติความสามารถ ปฏิสัมพันธ์ระหว่างวิธีการเชื่อมโยงคะแนนและโครงสร้างมิติความสามารถ และปฏิสัมพันธ์ระหว่างระดับความสัมพันธ์ระหว่างมิติความสามารถและโครงสร้างมิติความสามารถ มีผลทำให้ดัชนี Ln RMSE BIAS และ CORR ของค่าอำนาจจำแนกในมิติที่ 1 และ มิติที่ 2 ทั้ง 6 ตัวแปร โดยเฉลี่ยต่างกัน

สำหรับอิทธิพลหลัก พบว่า วิธีการเชื่อมโยงคะแนน ระดับความสัมพันธ์ระหว่างมิติความสามารถ และโครงสร้างมิติความสามารถที่ต่างกันทำให้มีค่าดัชนี Ln RMSE BIAS และค่าCORR ของค่าอำนาจจำแนกในมิติที่ 1 และ มิติที่ 2 โดยเฉลี่ยต่างกัน

เมื่อพิจารณาค่าสถิติจาก Levene's Test เพื่อศึกษาความแปรปรวนทั้ง 6 ตัวแปร ปรากฏดังตารางที่ 4.9

ตารางที่ 4.9 การทดสอบความแปรปรวนของค่าดัชนี RMSE BIAS และ CORR ของค่าอำนาจจำแนกในมิติที่ 1 และ มิติที่ 2 ทั้ง 6 ตัวแปร จาก Levene's Test

ตัวแปรตาม	Levene's Test			
	F	df1	df2	p
Ln RMSE_a1	1.185	11	228	.299
Ln RMSE_a2	1.012	11	228	.436
BIAS_a1	8.654	11	228	.000
BIAS_a2	13.695	11	228	.000
CORR_a1	8.116	11	228	.000
CORR_a2	7.352	11	228	.000

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

จากตารางที่ 4.9 การทดสอบความเท่ากันของความแปรปรวนด้วยวิธี Levene's Test เพื่อตรวจสอบความแปรปรวน ทั้ง 6 ตัวแปร จำแนกตามวิธีการเชื่อมโยงคะแนน ความสัมพันธ์ระหว่างมิติความสามารถ และระดับความสัมพันธ์ระหว่างมิติความสามารถ พบว่า มีลักษณะเหมือนการแปลงค่าพารามิเตอร์จากระดับที่ 1 ไปยังระดับที่ 2 นั่นคือ ความแปรปรวนของ Ln RMSE_a1 (F=1.185, p=.299) และ Ln RMSE_a2 (F=1.012, p=.436) จำแนกตามวิธีการเชื่อมโยงคะแนน ความสัมพันธ์ระหว่างมิติความสามารถ และระดับความสัมพันธ์ระหว่างมิติความสามารถ แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ดังนั้นผู้วิจัยจึงเลือกใช้การทดสอบภายหลัง (Post Hoc Tests) หรือการเปรียบเทียบพหุ (Multiple Comparison) เพื่อเปรียบเทียบความแตกต่างของค่าเฉลี่ยแต่ละกลุ่มที่เหมาะสมในกรณีที่ความแปรปรวนของตัวแปรตามในแต่ละกลุ่มมีค่าเท่ากันด้วย Sheffee ในขณะที่ความแปรปรวนของ BIAS_a_1 (F=8.654, p=.000), BIAS_a_2 (F=13.695, p=.000), CORR_a_1 (F=8.116, p=.000) และ CORR_a_2 (F=7.352, p=.000) แตกต่างกันอย่างมีนัยสำคัญทางสถิติ เมื่อใช้วิธีการเชื่อมโยงคะแนน ความสัมพันธ์ระหว่างมิติความสามารถ และระดับความสัมพันธ์ระหว่างมิติความสามารถต่างกัน ดังนั้นผู้วิจัยจึงเลือกใช้การทดสอบภายหลัง (Post Hoc Tests) หรือการเปรียบเทียบพหุ (Multiple Comparison) เพื่อเปรียบเทียบความแตกต่างของค่าเฉลี่ยแต่ละกลุ่มที่เหมาะสมในกรณีที่ความแปรปรวนของตัวแปรตามในแต่ละกลุ่มมีต่างกันด้วย Dunnett T3

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.10 การทดสอบอิทธิพลระหว่างตัวแปรอิสระ (Test of Between-subject Effects)

แหล่งความแปรปรวน	ตัวแปรตาม	Type III Sum of Squares	df	Mean Square	F	p	การเปรียบเทียบรายคู่
Corrected Model	1. Ln RMSE_a1	3.139	11	0.285	5.466	.000	
	2. Ln RMSE_a2	1.359	11	0.124	3.21	.000	
	3. BIAS_a1	0.616	11	0.056	80.06	.000	
	4. BIAS_a2	0.651	11	0.059	114.1	.000	
	5. CORR_a1	3.439	11	0.313	41.8	.000	
	6. CORR_a2	2.277	11	0.207	39.89	.000	
Intercept	1. Ln RMSE_a1	1027	1	1027	19673	.000	
	2. Ln RMSE_a2	1033	1	1033	26829	.000	
	3. BIAS_a1	3.568	1	3.568	5097	.000	
	4. BIAS_a2	3.470	1	3.47	6693	.000	
	5. CORR_a1	159.300	1	159.3	21305	.000	
	6. CORR_a2	159.100	1	159.1	30661	.000	
วิธีการเชื่อมโยงคะแนน (METHOD)	1. Ln RMSE_a1	0.139	1	0.139	2.661	.104	[1]M>NOP
	2. Ln RMSE_a2	0.144	1	0.144	3.741	.054	[1]M>NOP
	3. BIAS_a1	0.017	1	0.017	23.71	.000	M>NOP**
	4. BIAS_a2	0.008	1	0.008	16.32	.000	M>NOP**
	5. CORR_a1	0.002	1	0.002	0.236	.628	NOP>M
	6. CORR_a2	0.000	1	0.000	0.178	.674	NOP>M
โครงสร้างมิติความสามารถ (DS)	1. Ln RMSE_a1	0.311	1	0.311	5.964	.015	[1]MS>APSS*
	2. Ln RMSE_a2	0.064	1	0.064	1.661	.199	[1]MS>APSS
	3. BIAS_a1	0.072	1	0.072	102.300	.000	MS>APSS**
	4. BIAS_a2	0.018	1	0.018	34.810	.000	MS>APSS**
	5. CORR_a1	0.413	1	0.413	55.210	.000	APSS>MS**
	6. CORR_a2	0.301	1	0.301	58.010	.000	APSS>MS**
ระดับความสัมพันธ์ระหว่างมิติความสามารถ (r)	1. Ln RMSE_a1	1.513	2	0.757	14.490	.000	[1]0.8>0.0>0.5 [1]0.8>0.0**
	2. Ln RMSE_a2	0.931	2	0.465	12.090	.000	[1]0.8>0.0>0.5 [1]0.8>0.0**
	3. BIAS_a1	0.462	2	0.231	329.800	.000	0.0>0.5>0.8 0.5>0.8**
	4. BIAS_a2	0.598	2	0.299	576.700	.000	0.0>0.5>0.8**
	5. CORR_a1	2.889	2	1.444	193.100	.000	0.0>0.5>0.8**
	6. CORR_a2	1.945	2	0.972	187.400	.000	0.0>0.5>0.8**
METHOD X DS	1. Ln RMSE_a1	0.015	1	0.015	0.292	.590	
	2. Ln RMSE_a2	0.022	1	0.022	0.563	.454	
	3. BIAS_a1	0.004	1	0.004	6.000	.015	
	4. BIAS_a2	0.003	1	0.003	6.069	.014	
	5. CORR_a1	0.007	1	0.007	0.873	.351	
	6. CORR_a2	0.002	1	0.002	0.295	.587	

แหล่งความแปรปรวน	ตัวแปรตาม	Type III Sum of Squares	df	Mean Square	F	p	การเปรียบเทียบรายคู่
METHOD X r	1. Ln RMSE_a1	0.105	2	0.053	1.007	.367	
	2. Ln RMSE_a2	0.069	2	0.035	0.897	.409	
	3. BIAS_a1	0.006	2	0.003	4.502	.012	
	4. BIAS_a2	0.006	2	0.003	5.748	.004	
	5. CORR_a1	0.003	2	0.002	0.208	.812	
	6. CORR_a2	0.000	2	2E-04	0.042	.959	
DS X r	1. Ln RMSE_a1	1.031	2	0.516	9.877	.000	
	2. Ln RMSE_a2	0.121	2	0.061	1.575	.209	
	3. BIAS_a1	0.049	2	0.025	35.32	.000	
	4. BIAS_a2	0.013	2	0.007	12.92	.000	
	5. CORR_a1	0.117	2	0.058	7.819	.001	
	6. CORR_a2	0.028	2	0.014	2.673	.071	
METHOD X DS X r	1. Ln RMSE_a1	0.024	2	0.012	0.23	.795	
	2. Ln RMSE_a2	0.009	2	0.004	0.112	.894	
	3. BIAS_a1	0.007	2	0.003	4.683	.010	
	4. BIAS_a2	0.004	2	0.002	3.775	.024	
	5. CORR_a1	0.009	2	0.005	0.605	.547	
	6. CORR_a2	7E-04	2	0.000	0.071	.931	
Error	1. Ln RMSE_a1	11.9	228	0.052			
	2. Ln RMSE_a2	8.776	228	0.038			
	3. BIAS_a1	0.16	228	7E-04			
	4. BIAS_a2	0.118	228	5E-04			
	5. CORR_a1	1.705	228	0.007			
	6. CORR_a2	1.183	228	0.005			
Total	1. Ln RMSE_a1	1042	240				
	2. Ln RMSE_a2	1043	240				
	3. BIAS_a1	4.344	240				
	4. BIAS_a2	4.239	240				
	5. CORR_a1	164.5	240				
	6. CORR_a2	162.6	240				
Corrected Total	1. Ln RMSE_a1	15.04	239				
	2. Ln RMSE_a2	10.14	239				
	3. BIAS_a1	0.776	239				
	4. BIAS_a2	0.769	239				
	5. CORR_a1	5.144	239				
	6. CORR_a2	3.461	239				

*p< .05, **p<.01

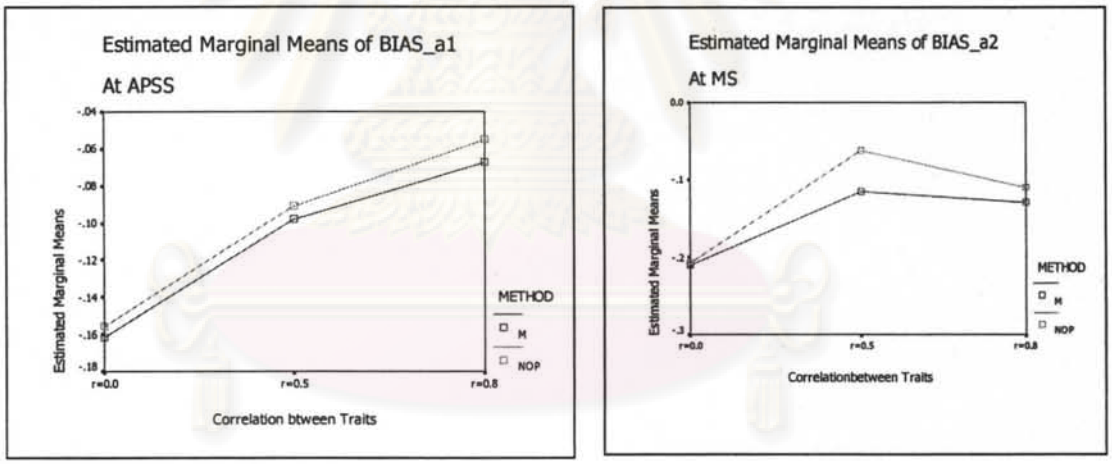
[1] เป็นการเปรียบเทียบความมากน้อยจากค่าเฉลี่ย RMSE

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

จากตารางที่ 4.10 สามารถสรุปผลการวิเคราะห์ความแปรปรวนพหุนาม ดังต่อไปนี้

1) ปฏิสัมพันธ์ระหว่าง 3 ตัวแปร (METHOD X DS X r)

ผลการวิเคราะห์ความแปรปรวนพหุนาม พบว่า เซ็นทรอยด์ของดัชนี BIAS ของอำนาจจำแนกในมิติที่ 1 และมิติที่ 2 คือ BIAS_a1 ($F_{(2,228)}$= 4.683 p=.010) และ BIAS_a2 ($F_{(2,228)}$= 3.775 p=.024) มีปฏิสัมพันธ์ระหว่างวิธีการเชื่อมโยงคะแนน (M) โครงสร้างมิติความสามารถ (DS) และระดับความสัมพันธ์ระหว่างมิติความสามารถ (r) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ในขณะที่เซ็นทรอยด์ของดัชนีบ่งชี้คุณภาพ Ln RMSE CORR ทั้งอำนาจจำแนกในมิติที่ 1 และมิติที่ 2 คือ Ln RMAE_a1 ($F_{(2,228)}$= 0.23 p=.795) Ln RMAE_a2 ($F_{(2,228)}$= .112 p=.894) CORR_a1 ($F_{(2,228)}$= .605 p=.547) และ CORR_a2 ($F_{(2,228)}$= .071 p=.931) แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ แสดงว่า วิธีการเชื่อมโยงคะแนน โครงสร้างมิติความสามารถ และระดับความสัมพันธ์ระหว่างมิติความสามารถ ไม่มีปฏิสัมพันธ์ที่ส่งผลต่อดัชนีบ่งชี้คุณภาพ RMSE และ CORR แต่มีปฏิสัมพันธ์ที่ส่งผลต่อดัชนี BIAS นั่นเอง โดยสามารถแสดงแผนภาพปฏิสัมพันธ์ของวิธีการเชื่อมโยงคะแนน ระดับความสัมพันธ์ระหว่างมิติความสามารถ ที่ระดับโครงสร้างมิติความสามารถต่างกัน ดังต่อไปนี้



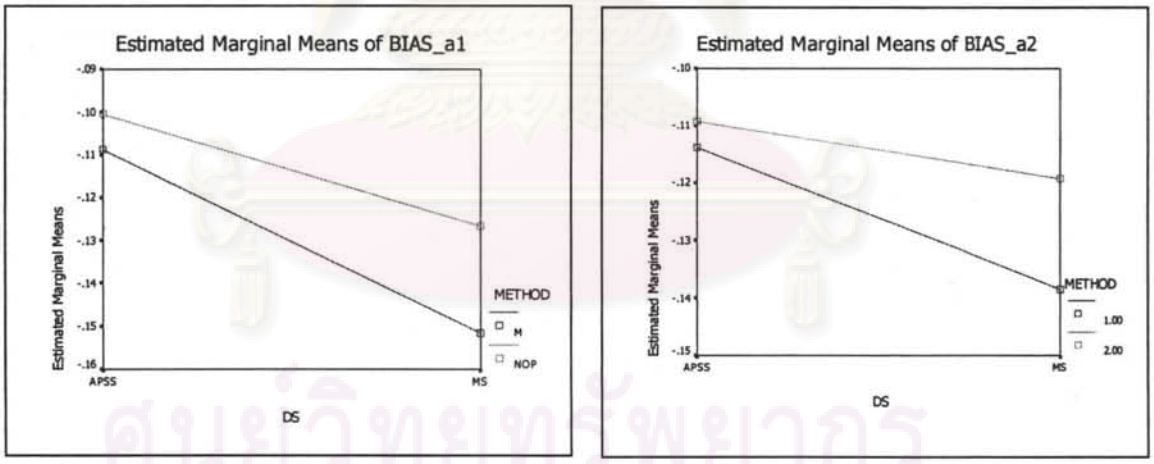
แผนภาพที่ 4.10 กราฟแสดงปฏิสัมพันธ์ระหว่างวิธีการเชื่อมโยงคะแนน ระดับความสัมพันธ์ระหว่างมิติความสามารถในโครงสร้างที่ไม่ซับซ้อน (APSS) และโครงสร้างที่ซับซ้อนมากขึ้น (MS) ซึ่งมีผลต่อค่าดัชนี BIAS (METHOD X r at DS)

จากแผนภาพแสดงให้เห็นได้อย่างชัดเจนว่า ในกรณีที่แบบสอบมีโครงสร้างที่ไม่ซับซ้อน (APSS) ผลของวิธีการเชื่อมโยงคะแนนและระดับความสัมพันธ์ระหว่างมิติความสามารถมีปฏิสัมพันธ์ร่วมกันในการส่งผลต่อค่าดัชนี BIAS ของค่าอำนาจจำแนกทั้ง 2 มิติ โดยในทุกระดับความสัมพันธ์วิธี M จะมีค่าดัชนี BIAS มากกว่าวิธี NOP ในทุกระดับความสัมพันธ์ แสดงให้เห็นว่า วิธี NOP มีการประมาณค่าใกล้เคียงกับค่าที่แท้จริงมากกว่าวิธี M ซึ่งสะท้อนให้เห็นถึงความถูกต้องในการเชื่อมโยงคะแนน

เมื่อพิจารณาในกรณีที่แบบสอบมีโครงสร้างที่ซับซ้อนมากขึ้น (MS) พบว่าผลของวิธีการเชื่อมโยงคะแนนและระดับความสัมพันธ์ระหว่างมิติความสามารถมีปฏิสัมพันธ์ร่วมกันในการส่งผลต่อค่า BIAS ของค่าอำนาจจำแนกทั้ง 2 มิติ โดยในทุกระดับความสัมพันธ์ วิธี M จะมีค่าดัชนี BIAS มากกว่าวิธี NOP แสดงให้เห็นว่า วิธี NOP มีการประมาณค่าใกล้เคียงกับค่าที่แท้จริงมากกว่าวิธี M โดยเมื่อระดับความสัมพันธ์อยู่ในระดับปานกลาง (r=0.5) จะมีค่า BIAS ต่ำกว่าระดับอื่น แสดงว่าเมื่อความสัมพันธ์อยู่ในระดับปานกลางจะมีการประมาณค่าใกล้เคียงกับค่าที่แท้จริงมากกว่าความสัมพันธ์ระหว่างมิติความสามารถในระดับอื่น

2) ปฏิสัมพันธ์ระหว่าง 2 ตัวแปร

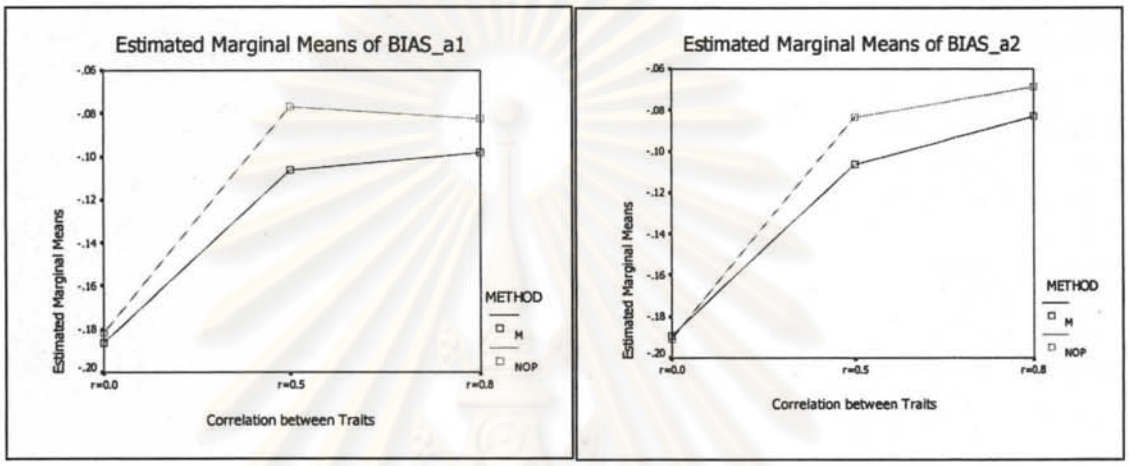
2.1) ปฏิสัมพันธ์ระหว่างวิธีการเชื่อมโยงคะแนนและโครงสร้างมิติความสามารถ (METHOD X DS) พบว่ามีลักษณะคล้ายคลึงกับการแปลงค่าพารามิเตอร์จากระดับที่ 1 ไปยังระดับที่ 2 นั่นคือ วิธีการเชื่อมโยงคะแนนและโครงสร้างมิติความสามารถมีปฏิสัมพันธ์ต่อค่าดัชนี BIAS ทั้งในมิติที่ 1 ($F_{(1,228)}$= 6.000, p=.015) และมิติที่ 2 ($F_{(1,228)}$= 6.069, p=.014) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ในขณะไม่มีปฏิสัมพันธ์ที่ส่งผลต่อค่าดัชนี RMSE และ CORR ทั้งในมิติที่ 1 และ มิติที่ 2 แสดงว่า วิธีการเชื่อมโยงคะแนน และโครงสร้างมิติความสามารถ ไม่มีปฏิสัมพันธ์ที่ส่งผลต่อดัชนีบ่งชี้คุณภาพ RMSE และ CORR แต่มีปฏิสัมพันธ์กันที่ส่งผลต่อค่าดัชนี BIAS นั่นเอง โดยสามารถแสดงแผนภาพปฏิสัมพันธ์ระหว่าง 2 ตัวแปร ดังต่อไปนี้



แผนภาพที่ 4.11 กราฟแสดงปฏิสัมพันธ์ METHOD X DS

จากแผนภาพ จะเห็นว่า เส้นกราฟของวิธี M และ วิธี NOP ไม่ขนานกัน จากภาพไม่ว่าจะเป็นโครงสร้างในลักษณะใด วิธี NOP จะมีค่า BIAS ที่ต่ำกว่าวิธี M แสดงว่าวิธี NOP มีการประมาณค่าได้ใกล้เคียงกับค่าที่แท้จริงมากกว่าวิธี M

2.2) ปฏิสัมพันธ์ระหว่างวิธีการเชื่อมโยงคะแนนและระดับความสัมพันธ์ระหว่างมิติความสามารถ (METHOD X r) พบว่า วิธีการเชื่อมโยงคะแนนและโครงสร้างมิติความสามารถมีปฏิสัมพันธ์ต่อค่าดัชนี BIAS ทั้งในมิติที่ 1 ($F_{(2,228)}$= 4.502, p=.012) และมิติที่ 1 ($F_{(2,228)}$= 5.748, p=.004) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่า วิธีการเชื่อมโยงคะแนน และโครงสร้างมิติความสามารถ มีปฏิสัมพันธ์ที่ส่งผลต่อดัชนี BIAS นั่นเอง สามารถแสดงแผนภาพปฏิสัมพันธ์ระหว่าง 2 ตัวแปร ดังต่อไปนี้



แผนภาพที่ 4.12 กราฟแสดงปฏิสัมพันธ์ METHOD X r

จากแผนภาพ จะเห็นว่า เส้นกราฟของวิธี M และ วิธี NOP ไม่ขนานกัน จากภาพไม่ว่าจะเป็นความสัมพันธ์ระหว่างมิติความสามารถในระดับใด วิธี NOP จะมีค่า BIAS ที่ต่ำกว่าวิธี M แสดงว่าวิธี NOP มีการประมาณค่าได้ใกล้เคียงกับค่าที่แท้จริงมากกว่าวิธี M

2.3) ปฏิสัมพันธ์ระหว่างโครงสร้างมิติความสามารถและระดับความสัมพันธ์ระหว่างมิติความสามารถ (DSX r) พบว่า โครงสร้างมิติความสามารถและระดับความสัมพันธ์ระหว่างมิติความสามารถมีปฏิสัมพันธ์ที่ส่งผลต่อค่าดัชนี BIAS และ CORR ในมิติที่ 1 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แสดงให้เห็นว่าโครงสร้างมิติความสามารถและระดับความสัมพันธ์ระหว่างมิติความสามารถ มีปฏิสัมพันธ์ที่ส่งผลต่อดัชนี BIAS และ CORR แต่ไม่ส่งผลต่อดัชนี RMSE

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

3) ผลของอิทธิพลหลัก

3.1) เมื่อพิจารณาวิธีการเชื่อมโยงคะแนน พบว่ามีอิทธิพลต่อค่าดัชนี BIAS ทั้งในมิติที่ 1 และมิติที่ 2 อย่างมีนัยสำคัญทางสถิติ แต่ไม่มีอิทธิพลต่อดัชนี RMSE และ CORR โดยเมื่อทำการเปรียบเทียบความแตกต่างของดัชนีบ่งชี้คุณภาพ ทั้งในมิติที่ 1 และ 2 พบว่า วิธี NOP มีการประมาณค่าได้ใกล้เคียงกับค่าที่แท้จริงมากกว่าวิธี M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ซึ่งพิจารณาจากดัชนี BIAS แต่เมื่อพิจารณาถึงดัชนี RMSE ซึ่งสะท้อนถึงความคงที่ของการเชื่อมโยงคะแนน และค่า CORR พบว่าทั้ง 2 วิธี ไม่แตกต่างกัน

3.2) สำหรับโครงสร้างมิติความสามารถพบว่ามีอิทธิพลต่อค่าดัชนี RMSE BIAS และ CORR ทั้งในมิติที่ 1 และมิติที่ 2 โดยเมื่อทำการเปรียบเทียบความแตกต่างของดัชนีบ่งชี้คุณภาพทั้งในมิติที่ 1 และ 2 พบว่า โครงสร้างที่ไม่ซับซ้อน (APSS) มีความคงที่ในการประมาณค่าและการประมาณค่าใกล้เคียงกับค่าที่แท้จริงมากกว่าโครงสร้างที่ซับซ้อนมากขึ้น อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ซึ่งพิจารณาจากดัชนี Ln RMSE และ BIAS ตามลำดับ และเมื่อพิจารณาถึงค่า CORR พบว่าโครงสร้างที่ไม่ซับซ้อน (APSS) มีค่า CORR สูงกว่าโครงสร้างที่ซับซ้อนมากขึ้น (MS) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

3.3) ส่วนระดับความสัมพันธ์ระหว่างมิติความสามารถ พบว่า มีอิทธิพลต่อค่าดัชนีบ่งชี้คุณภาพทั้งในมิติที่ 1 และมิติที่ 2 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยเมื่อทำการเปรียบเทียบความแตกต่างของดัชนี RMSE ทั้งในมิติที่ 1 และ 2 พบว่า เมื่อมีระดับความสัมพันธ์เท่ากับ 0.5 จะมีความคงที่ในการประมาณค่าที่ดีกว่าระดับความสัมพันธ์ 0.0 และ 0.8 ตามลำดับ สำหรับความถูกต้องในการประมาณค่า ซึ่งพิจารณาจากดัชนี BIAS พบว่า เมื่อมีระดับความสัมพันธ์เท่ากับ 0.8 จะมีการประมาณค่าใกล้เคียงกับค่าที่แท้จริงมากกว่าระดับความสัมพันธ์ 0.5 และ 0.0 ตามลำดับ และเมื่อพิจารณาถึงค่า CORR พบว่าเมื่อความสัมพันธ์มีระดับสูงขึ้นจะมีค่าลดลงอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

จากที่กล่าวมาเป็นการวิเคราะห์ความแปรปรวนของค่าเฉลี่ยดัชนีบ่งชี้คุณภาพ 3 ชนิด คือ RMSE BIAS และ CORR ซึ่งถือเป็นตัวแปรตามของการวิจัยในครั้งนี้ โดยมุ่งเน้นเปรียบเทียบระหว่างวิธีการเชื่อมโยงคะแนน 2 วิธี คือวิธี M กับวิธี NOP กับเงื่อนไขที่ศึกษา คือ โครงสร้างมิติความสามารถ และระดับความสัมพันธ์ระหว่างมิติความสามารถ สามารถสรุปผลการศึกษาปฏิสัมพันธ์ และผลการเปรียบเทียบรายคู่ในแต่ละตัวแปรต้นซึ่งถือเป็นอิทธิพลหลักของการวิจัย ได้ดังตารางที่ 4.11

ตารางที่ 4.11 สรุปผลการทดสอบอิทธิพลระหว่างตัวแปรอิสระ (Test of Between-subject Effects) ของการแปลงคะแนนจากระดับ 1 ไปยังระดับ 2 (L1_2) และ การแปลงคะแนนจากระดับ 2 ไปยังระดับ 3 (L2_3)

แหล่งความแปรปรวน	ตัวแปรตาม	L1_2		L2_3	
		F	เปรียบเทียบรายคู่	F	เปรียบเทียบรายคู่
วิธีการเชื่อมโยงคะแนน(METHOD)	1. Ln RMSE_a1	5.249*	M>NOP*	2.661	M>NOP
	2. Ln RMSE_a2	11.414**	M>NOP**	3.741	M>NOP
	3. BIAS_a1	45.643**	M>NOP**	23.710**	M>NOP**
	4. BIAS_a2	32.003**	M>NOP**	16.320**	M>NOP**
	5. CORR_a1	1.757	NOP>M	0.236	NOP>M
	6. CORR_a2	1.38	NOP>M	0.178	NOP>M
โครงสร้างมิติความสามารถ (DS)	1. Ln RMSE_a1	2.116	MS>APSS	5.964*	MS>APSS*
	2. Ln RMSE_a2	41.429**	MS>APSS**	1.661	MS>APSS
	3. BIAS_a1	13.356**	MS>APSS**	102.300**	MS>APSS**
	4. BIAS_a2	3.248	MS>APSS	34.810**	MS>APSS**
	5. CORR_a1	60.085**	APSS>MS**	55.210**	MS>APSS**
	6. CORR_a2	150.687**	APSS>MS**	58.010**	MS>APSS**
ระดับความสัมพันธ์ระหว่างมิติความสามารถ (r)	1. Ln RMSE_a1	26.523**	0.8>0.0>0.5**	14.490**	0.8>0.0>0.5 0.8>0.0**
	2. Ln RMSE_a2	56.275**	0.8>0.0>0.5**	12.090**	0.8>0.0>0.5 0.8>0.0**
	3. BIAS_a1	98.223**	0.0>0.5>0.8**	329.800**	0.0>0.5>0.8 0.5>0.8**
	4. BIAS_a2	106.046**	0.0>0.5>0.8**	576.700**	0.0>0.5>0.8**
	5. CORR_a1	300.075**	0.0>0.5>0.8**	193.100**	0.0>0.5>0.8**
	6. CORR_a2	233.218**	0.0>0.5>0.8**	187.400**	0.0>0.5>0.8**
METHOD X DS	1. Ln RMSE_a1	1.383	-	0.292	-
	2. Ln RMSE_a2	2.226	-	0.563	-
	3. BIAS_a1	31.303**	-	6.000*	-
	4. BIAS_a2	13.602**	-	6.069*	-
	5. CORR_a1	2.944	-	0.873	-
	6. CORR_a2	3.172	-	0.295	-
METHOD X r	1. Ln RMSE_a1	2.325	-	1.007	-
	2. Ln RMSE_a2	4.108*	-	0.897	-
	3. BIAS_a1	3.848*	-	4.502*	-
	4. BIAS_a2	1.965	-	5.748**	-
	5. CORR_a1	1.386	-	0.208	-
	6. CORR_a2	2.346	-	0.042	-
DS X r	1. Ln RMSE_a1	7.014**	-	9.877**	-
	2. Ln RMSE_a2	8.899**	-	1.575	-
	3. BIAS_a1	6.669**	-	35.32**	-
	4. BIAS_a2	0.103	-	12.92**	-
	5. CORR_a1	6.163**	-	7.819**	-
	6. CORR_a2	8.266**	-	2.673	-
METHOD XDS X r	1. Ln RMSE_a1	0.766	-	0.23	-
	2. Ln RMSE_a2	0.776	-	0.112	-
	3. BIAS_a1	0.356	-	4.683*	-
	4. BIAS_a2	0.374	-	3.775*	-
	5. CORR_a1	3.573*	-	0.605	-
	6. CORR_a2	3.257*	-	0.071	-

*p< .05, **p<.01

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

จากตารางที่ 4.11 เมื่อพิจารณาผลที่ได้จากการแปลงคะแนนจากระดับที่ 1 ไปยังระดับที่ 2 พบว่า วิธีการเชื่อมโยงคะแนน โครงสร้างมิติความสามารถ และระดับความสัมพันธ์ระหว่างมิติความสามารถ มีปฏิสัมพันธ์ที่ส่งผลต่อค่าความสัมพันธ์ระหว่างคะแนนที่แปลงได้กับคะแนนเกณฑ์ แต่ไม่มีปฏิสัมพันธ์ที่ส่งผลต่อความคงที่และความถูกต้องของการเชื่อมโยงคะแนน แต่เมื่อพิจารณาผลที่ได้จากการแปลงคะแนนจากระดับที่ 2 ไปยังระดับที่ 3 พบว่า วิธีการเชื่อมโยงคะแนน โครงสร้างมิติความสามารถ และระดับความสัมพันธ์ระหว่างมิติความสามารถ มีปฏิสัมพันธ์ที่ส่งผลต่อความถูกต้องของการเชื่อมโยงคะแนน แต่ไม่มีปฏิสัมพันธ์ที่ส่งผลต่อความคงที่ของการเชื่อมโยงคะแนนและไม่ส่งผลต่อค่าความสัมพันธ์ระหว่างคะแนนที่แปลงได้กับคะแนนเกณฑ์

เมื่อเปรียบเทียบค่าเฉลี่ยของดัชนีบ่งชี้คุณภาพระหว่างวิธีการเชื่อมโยงคะแนน 2 วิธี โดยภาพรวมพบว่า วิธี NOP มีความคงที่และมีความถูกต้องในการประมาณค่ามากกว่าวิธี M แต่ความสัมพันธ์ระหว่างค่าที่ได้จากการแปลงคะแนนกับคะแนนเกณฑ์ทั้งสองวิธีแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

เมื่อเปรียบเทียบค่าเฉลี่ยของดัชนีบ่งชี้คุณภาพระหว่างโครงสร้างมิติความสามารถระหว่างโครงสร้างที่ไม่ซับซ้อน (APSS) และโครงสร้างที่ซับซ้อนมากขึ้น (MS) โดยภาพรวม พบว่า โครงสร้างที่ไม่ซับซ้อน (APSS) มีความคงที่และมีความถูกต้องในการประมาณค่ามากกว่าโครงสร้างที่ซับซ้อนมากขึ้น (MS) โดยโครงสร้างที่ไม่ซับซ้อน (APSS) มีความสัมพันธ์ระหว่างค่าที่ได้จากจากการแปลงคะแนนกับคะแนนเกณฑ์ (CORR) สูงกว่าโครงสร้างที่ซับซ้อนมากขึ้น (MS)

เมื่อเปรียบเทียบค่าเฉลี่ยของดัชนีบ่งชี้คุณภาพระหว่างระดับความสัมพันธ์ระหว่างมิติความสามารถ (r) ในภาพรวมพบว่า เมื่อความสัมพันธ์เท่ากับ 0.5 จะมีความคงที่ในการประมาณค่าที่ดีกว่าระดับความสัมพันธ์ 0.0 และ 0.8 ตามลำดับ เมื่อพิจารณาถึงความถูกต้องในการประมาณค่า พบว่า เมื่อระดับความสัมพันธ์เท่ากับ 0.8 จะมีการประมาณค่าใกล้เคียงกับค่าที่แท้จริงมากกว่าระดับความสัมพันธ์ 0.5 และ 0.0 ตามลำดับ และเมื่อพิจารณาถึงค่า CORR ที่สะท้อนถึงความสัมพันธ์ระหว่างค่าที่ได้จากจากการแปลงคะแนนกับคะแนนเกณฑ์ พบว่าเมื่อความสัมพันธ์มีระดับสูงขึ้นจะมีค่า CORR ลดลง

รายละเอียดต่อไปเป็นการเปรียบเทียบค่าดัชนีบ่งชี้คุณภาพ RMSE BIAS และ CORR รายสถานการณ์ โดยมุ่งเน้นเปรียบเทียบระหว่างวิธีการเชื่อมโยงคะแนนกับเงื่อนไขที่ศึกษา ในการเปรียบเทียบระหว่างวิธี M และ NOP พร้อมทั้งนำเสนอการกระจายค่าดัชนีบ่งชี้คุณภาพจากการทำซ้ำเงื่อนไขละ 20 ครั้ง โดยนำเสนอเป็นแผนภาพ Box Plot ซึ่งมีรายละเอียดดังนี้

ตอนที่ 2 ผลการเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนนรายสถานการณ์

สำหรับผลการเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนนตามโมเดล MIRT ระหว่างวิธี M และวิธี NOP ภายใต้โครงสร้างมิติความสามารถและระดับความสัมพันธ์ระหว่างมิติความสามารถที่ต่างกัน โดยพิจารณาจากการเปรียบเทียบดัชนี RMSE BIAS และ CORR โดยผลการวิจัยพบว่า ผลที่ได้จากการแปลงค่าพารามิเตอร์ความยากของข้อสอบ (d) ให้คุณภาพของการเชื่อมโยงคะแนนไม่แตกต่างกัน แต่ผลที่ได้จากการแปลงค่าพารามิเตอร์อำนาจจำแนกแตกต่างกันตามเงื่อนไขที่ศึกษา ดังนั้นจึงมุ่งเน้นเฉพาะการเปรียบเทียบค่าเฉลี่ยของดัชนีคุณภาพอำนาจจำแนกของข้อสอบ (a_1 และ a_2) สามารถสรุปผลการเปรียบเทียบเป็น 2 ลักษณะ

1) ผลการเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนนระหว่างวิธี M กับวิธี NOP ของแต่ละโครงสร้างมิติความสามารถ (APSS และ MS) และแต่ละระดับความสัมพันธ์มิติความสามารถ (r=0.0, r=0.5 และ r=0.8) ดังนั้นการวิเคราะห์ส่วนนี้ตัวแปรอิสระของการวิจัยคือวิธีการเชื่อมโยงคะแนนในแต่ละเงื่อนไขที่ศึกษา ส่วนตัวแปรตามคือคุณภาพของการเชื่อมโยงคะแนนประกอบด้วย 6 ตัวแปร คือ 1) Ln RMSE_a1 2) Ln RMSE_a2 3) BIAS_a1 4) BIAS_a2 5) CORR_a_1 และ 6) CORR_a_2 ดังนั้นในการเปรียบเทียบความแตกต่างระหว่างค่าเฉลี่ยจึงทำการวิเคราะห์ความแปรปรวนแบบพหุจำแนก 1 ทาง (One-way MANOVA) เพื่อให้ทราบว่าในแต่ละโครงสร้างมิติความสามารถ (APSS และ MS) และแต่ละระดับความสัมพันธ์มิติความสามารถ (r=0.0, r=0.5 และ r=0.8) วิธี M หรือวิธี NOP มีคุณภาพของการเชื่อมโยงคะแนนแตกต่างกันหรือไม่อย่างไร

2) ผลการเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนนระหว่างวิธี M กับวิธี NOP ภายใต้ 6 เงื่อนไข คือ APSS_0.0, MS_0.0, APSS_0.0, MS_0.5, APSS_0.8 และ MS_0.8 โดยใช้สถิติทดสอบ t พร้อมทั้งนำเสนอแผนภาพในรูป BOX Plot เพื่อศึกษาลักษณะของการกระจายข้อมูลที่มีการทำซ้ำ 20 ครั้ง ทั้ง 6 เงื่อนไข ตามรายดัชนีบ่งชี้คุณภาพทั้ง 3 ชนิด คือ RMSE BIAS และ CORR ทั้งในมิติที่ 1 และ มิติที่ 2 เพื่อให้ทราบว่าเมื่อศึกษา วิธี M หรือวิธี NOP มีคุณภาพของการเชื่อมโยงคะแนนแตกต่างกันหรือไม่อย่างไรในแต่ละเงื่อนไขที่มีลักษณะเป็น Nested Design

ในตอนนี้แบ่งการนำเสนอแยกแต่ละระดับของการเชื่อมโยงคะแนน คือระดับที่ 1 ไปยังระดับที่ 2 (L1_2) และระดับที่ 2 ไปยังระดับที่ 3 (L2_3) มีรายละเอียดต่อไปนี้

2.1 ผลการเปรียบเทียบค่าเฉลี่ยของดัชนี RMSE BIAS และค่า CORR จากระดับที่ 1 ไปยังระดับที่ 2 รายสถานการณ์

2.1.1 ผลการเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนนระหว่างวิธี M กับวิธี NOP ของแต่ละโครงสร้างมิติความสามารถ (APSS และ MS) และแต่ละระดับความสัมพันธ์มิติความสามารถ (r=0.0, r=0.5 และ r=0.8) โดยใช้สถิติทดสอบ One-way MANOVA ได้ผลดังตารางที่ 4.12

ตารางที่ 4.12 ผลการทดสอบความแปรปรปรวนแบบพหุจำแนกทางเดียว (One-way MANOVA) เพื่อเปรียบเทียบค่าเฉลี่ยของดัชนี Ln RMSE BIAS และค่า CORR ระหว่างวิธี M และวิธี NOP ในแต่ละเงื่อนไขที่ศึกษา จากระดับที่ 1 ไปยังระดับที่ 2 (L1_2)

เงื่อนไขการจำลองข้อมูล		วิธีการ	Ln RMSE				BIAS				CORR			
			a1		a2		a1		a2		a1		a2	
			M (SD)	F	M (SD)	F	M (SD)	F	M (SD)	F	M (SD)	F	M (SD)	F
DS	APSS	M	-1.924 (0.213)	0.480	-2.010 (0.183)	1.493	-0.084 (0.031)	0.652	-0.086 (0.037)	1.651	0.814 (0.120)	0.028	0.837 (0.099)	0.095
		NOP	-1.950 (0.203)		-2.051 (0.181)		-0.079 (0.036)		-0.077 (0.040)		0.810 (0.131)		0.831 (0.109)	
	MS	M	-1.862 (0.227)	**4.772***	-1.839 (0.246)	**6.152***	-0.124 (0.043)	27.768	-0.112 (0.050)	**16.816****	0.720 (0.181)	1.027	0.700 (0.174)	1.043
		NOP	-1.944 (0.181)		-1.944 (0.216)		-0.071 (0.065)		-0.069 (0.064)		0.751 (0.141)		0.729 (0.130)	
r	r=0.0	M	-1.896 (0.214)	0.122	-1.913 (0.176)	0.119	-0.134 (0.042)	1.888	-0.132 (0.028)	**5.100***	0.884 (0.047)	0.368	0.869 (0.063)	0.016
		NOP	-1.912 (0.199)		-1.927 (0.179)		-0.121 (0.041)		-0.117 (0.032)		0.890 (0.042)		0.868 (0.067)	
	r=0.5	M	-2.022 (0.167)	0.312	-2.094 (0.152)	1.829	-0.103 (0.022)	**29.090****	-0.103 (0.021)	**28.587****	0.830 (0.059)	0.036	0.828 (0.065)	0.116
		NOP	-2.042 (0.155)		-2.140 (0.156)		-0.070 (0.032)		-0.077 (0.022)		0.827 (0.063)		0.823 (0.067)	
	r=0.8	M	-1.760 (0.201)	**8.503****	-1.767 (0.237)	**10.280****	-0.075 (0.038)	**21.107****	-0.062 (0.051)	**10.596****	0.588 (0.147)	1.460	0.608 (0.161)	1.626
		NOP	-1.886 (0.185)		-1.925 (0.203)		-0.034 (0.041)		-0.024 (0.051)		0.624 (0.116)		0.649 (0.124)	

*p< .05, **p<.01

ตารางที่ 4.13 ผลการทดสอบรายคู่ที่มีนัยสำคัญทางสถิติที่ระดับ .05

เงื่อนไขที่ศึกษา		RMSE		BIAS		CORR	
		a_1	a_2	a_1	a_2	a_1	a_2
โครงสร้างมิติความสามารถ (DS)	APSS	X	X	X	X	X	X
	MS	**M>NOP**	**M>NOP**	X	**M>NOP**	X	X
ระดับความสัมพันธ์ระหว่างมิติความสามารถ (r)	r=0.0	X	X	X	**M>NOP**	X	X
	r=0.5	X	X	**M>NOP**	**M>NOP**	X	X
	r=0.8	**M>NOP**	**M>NOP**	**M>NOP**	**M>NOP**	X	X

X ไม่พบว่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติ

จากตารางที่ 4.12 และ 4.13 สามารถสรุปได้การเชื่อมโยงคะแนนสำหรับชุดของข้อสอบที่มีโครงสร้างไม่ซับซ้อน (APSS) พบว่าทั้ง 2 วิธี มีคุณภาพของการเชื่อมโยงคะแนนไม่แตกต่างกัน แต่เมื่อโครงสร้างมีความซับซ้อนมากขึ้น (MS) พบว่าวิธี M มีค่าดัชนี RMSE และ BIAS มากกว่าวิธี NOP สะท้อนให้เห็นว่าวิธี NOP มีความคงที่และมีความถูกต้องของการเชื่อมโยงคะแนนมากกว่าวิธี M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

เมื่อพิจารณาระดับความสัมพันธ์ระดับมิติความสามารถ พบว่า ในทุกระดับความสัมพันธ์ระหว่างมิติความสามารถวิธี NOP จะมค่า BIAS ต่ำกว่าวิธี M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 สะท้อนให้เห็นว่าวิธี NOP มีความถูกต้องในการประมาณค่ามากกว่าวิธี M อย่างไรก็ตามจากผลการวิจัยพบว่าระดับความสัมพันธ์ระหว่างมิติความสามารถจะส่งผลกระทบน้อยมากต่อค่า RMSE ซึ่งสะท้อนถึงความคงที่ของการเชื่อมโยงคะแนน อย่างไรก็ตามพบว่าเมื่อความสัมพันธ์ระหว่างมิติความสามารถสูง (r=0.8) วิธี NOP จะมีค่า RMSE ต่ำกว่าวิธี M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าวิธี NOP มีความคงที่ในการเชื่อมโยงคะแนนมากกว่าวิธี M

ข้อสังเกตที่ได้จากผลการวิจัยครั้งนี้พบว่า ค่า CORR ทั้ง 2 วิธี ไม่ต่างกันไม่ว่าจะศึกษาจากโครงสร้างมิติความสามารถลักษณะใด และ ระดับความสัมพันธ์ระหว่างมิติความสามารถระดับใด แสดงว่าทั้งสองวิธีมีความสัมพันธ์ระหว่างค่าที่แปลงได้จากแบบสอบปรับเทียบคะแนนกับคะแนนเกณฑ์ที่ได้จากแบบสอบฐานไม่ต่างกัน

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

2.1.2 ผลการเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนนระหว่างวิธี M กับวิธี NOP ภายใต้ 6 เงื่อนไข คือ APSS_0.0, MS_0.0, APSS_0.0, MS_0.5, APSS_0.8 และ MS_0.8 โดยใช้สถิติทดสอบ t ปรากฏผลดังตารางที่ 4.14

ตารางที่ 4.14 การเปรียบเทียบค่าเฉลี่ยของดัชนี RMSE BIAS และค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างวิธี M และวิธี NOP จากระดับที่ 1 ไปยังระดับที่ 2 (L1_2)

เงื่อนไขการจำลองข้อมูล	วิธีการ	RMSE				BIAS				Correlation Coefficient			
		a1		a2		a1		a2		a1		a2	
		M (SD)	t	M (SD)	t	M (SD)	t	M (SD)	t	M (SD)	t	M (SD)	t
Corr. 0.0 APSS	M	0.157	0.277	0.143	0.147	-0.103	1.310	-0.121	-0.359	0.906	-1.019	0.917	-0.023
		(0.033)		(0.022)		(0.033)		(0.023)		(0.027)		(0.020)	
(APSS_0.0)	NOP	0.154		0.142		-0.115		-0.118		0.914		0.918	
		(0.029)		(0.021)		(0.028)		(0.021)		(0.025)		(0.021)	
Corr. 0.0 MS	M	0.151	0.283	0.157	0.337	-0.166	**-3.11****	-0.144	**-2.520****	0.862	-0.253	0.822	0.213
		(0.038)		(0.028)		(0.021)		(0.027)		(0.054)		(0.054)	
(MS_0.0)	NOP	0.147		0.154		-0.128		-0.117		0.865		0.818	
		(0.034)		(0.028)		(0.050)		(0.041)		(0.042)		(0.059)	
Corr. 0.5 APSS	M	0.135	0.154	0.12	0.546	-0.080	**-3.238****	-0.091	**-2.069***	0.858	0.083	0.872	0.426
		(0.027)		(0.019)		(0.016)		(0.016)		(0.051)		(0.047)	
(APSS_0.5)	NOP	0.133		0.116		-0.077		-0.081		0.857		0.865	
		(0.026)		(0.018)		(0.011)		(0.012)		(0.055)		(0.051)	
Corr. 0.5 MS	M	0.134	0.819	0.13	1.447	-0.115	**-4.840****	-0.115	**-5.56****	0.801	0.219	0.785	0.209
		(0.020)		(0.017)		(0.021)		(0.018)		(0.054)		(0.051)	
(MS_0.5)	NOP	0.129		0.122		-0.062		-0.073		0.797		0.782	
		(0.015)		(0.018)		(0.043)		(0.028)		(0.056)		(0.053)	
Corr. 0.8 APSS	M	0.157	0.854	0.147	1.518	-0.059	-1.706	-0.048	-1.994	0.678	0.544	0.721	0.405
		(0.032)		(0.026)		(0.025)		(0.025)		(0.103)		(0.075)	
(APSS_0.8)	NOP	0.148		0.134		-0.046		-0.033		0.659		0.71	
		(0.031)		(0.025)		(0.024)		(0.023)		(0.111)		(0.098)	
Corr. 0.8 MS	M	0.194	**4.115****	0.204	**4.328****	-0.091	**4.582****	-0.076	**-2.791****	0.499	**-2.348***	0.495	**-2.23***
		(0.025)		(0.032)		(0.042)		(0.066)		(0.129)		(0.144)	
(MS_0.8)	NOP	0.160		0.163		-0.023		-0.016		0.589		0.588	
		(0.026)		(0.028)		(0.052)		(0.069)		(0.113)		(0.119)	

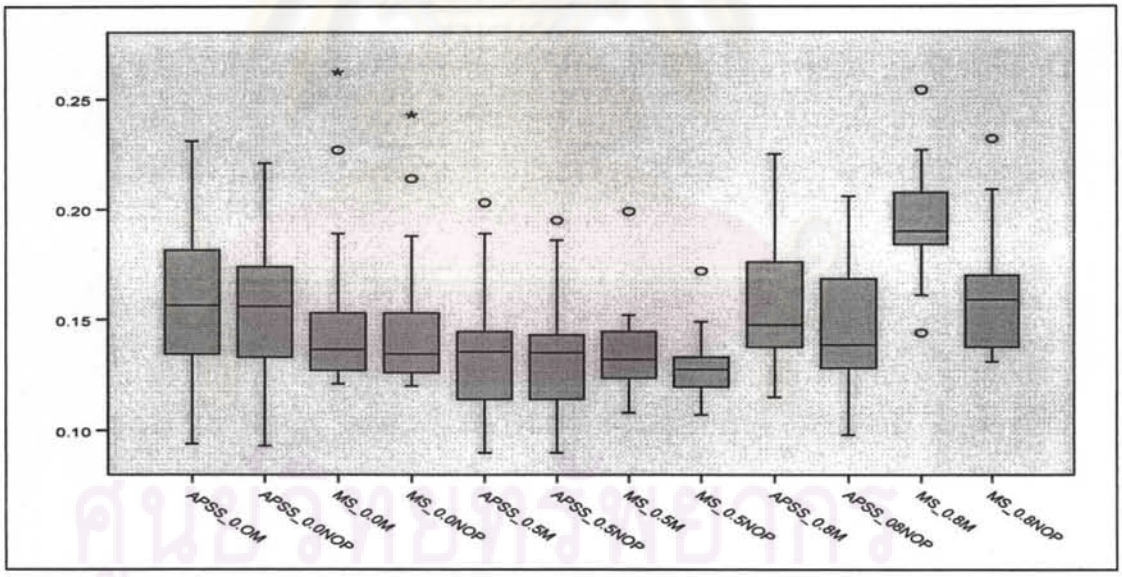
*p< .05, **p<.01

จากตารางที่ 4.14 ผลการเปรียบเทียบค่าเฉลี่ยของดัชนีบ่งชี้คุณภาพแต่ละชนิดในแต่ละระดับของตัวแปรเงื่อนไขของการจำลองข้อมูล ได้ผลดังต่อไปนี้

2.1.1 ผลการวิเคราะห์โดยใช้ดัชนี Ln RMSE เป็นตัวแปรตาม พบว่ามีเพียงเงื่อนไขของความสัมพันธ์ระหว่างมิติความสามารถอยู่ในระดับสูงที่มีโครงสร้างซับซ้อน (MS_0.8) เท่านั้น ที่มีผลทำให้ค่าดัชนี RMSE ของวิธี M แตกต่างจากวิธี NOP อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ทั้งค่าพารามิเตอร์ a_1 (t=4.115) และ a_2 (t=4.328) โดยวิธี NOP (RMSE_a_1=0.160, RMSE_a_2=0.163) มีความคงที่ของการประมาณค่าสูงกว่าวิธี M (RMSE_a_1=0.194, RMSE_a_2=0.204) อย่างเห็นได้ชัด สำหรับเงื่อนไขอื่นๆ พบว่ามีค่าเฉลี่ยของดัชนี Ln RMSE แตกต่างกันแต่ไม่มีนัยสำคัญทางสถิติ อย่างไรก็ตามแม้ผลการวิเคราะห์พบว่าทั้ง 2 วิธีให้ผลไม่แตกต่างกันในแต่ละเงื่อนไข แต่ข้อค้นพบในการศึกษาจากค่าเฉลี่ยของดัชนี RMSE สำหรับค่าพารามิเตอร์ a_1 และ a_2 พบว่า ความสัมพันธ์ระหว่างมิติอยู่ในระดับปานกลาง (r=0.5) มีความคงที่ของการประมาณค่ามากกว่าความสัมพันธ์ระดับอื่น

อย่างไรก็ตามเมื่อความสัมพันธ์ระหว่างมิติความสามารถในระดับสูง (r=0.8) พบว่ามีความคลาดเคลื่อนของการประมาณค่ามากกว่าระดับความสัมพันธ์อื่นๆ จากที่กล่าวมาสามารถแสดงแผนภาพ Box Plot เพื่อพิจารณาลักษณะการกระจายของค่าความคลาดเคลื่อน RMSE ของค่าพารามิเตอร์การแปลงคะแนนในมิติที่ 1 และ 2 ดังต่อไปนี้

1) ดัชนี RMSE สำหรับค่าอำนาจจำแนกของข้อสอบในมิติที่ 1 (a1)

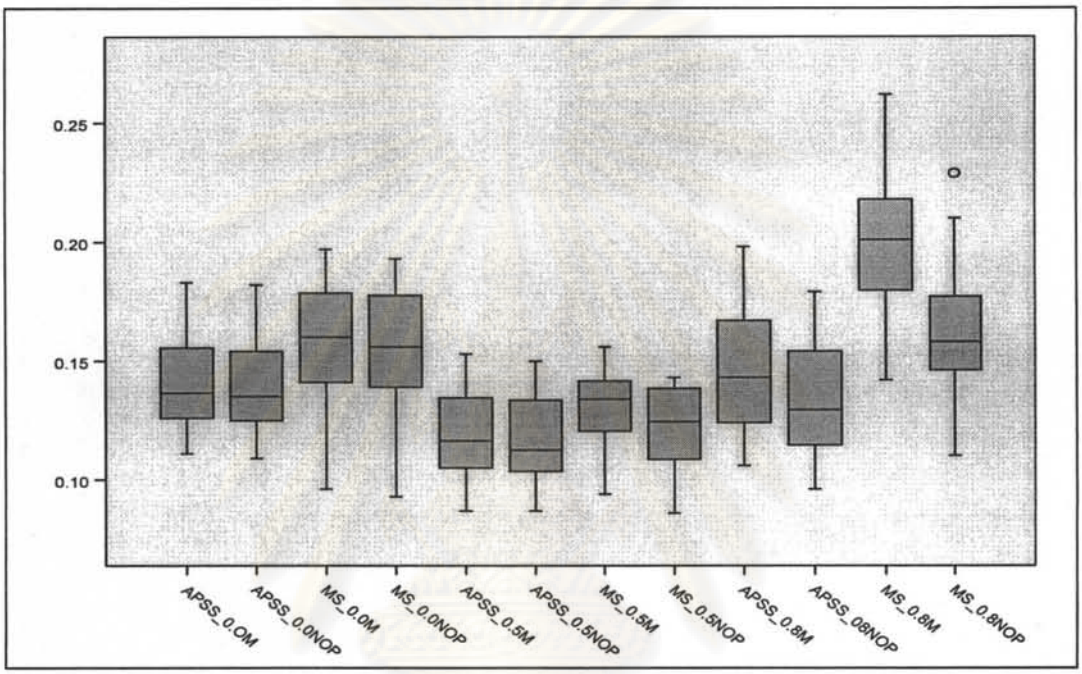


แผนภาพที่ 4.13 แผนภูมิ Box Plot ของข้อมูลชุดวัดซ้ำทั้งหมด 20 ครั้ง สำหรับดัชนี RMSE ของค่าอำนาจจำแนกของข้อสอบในมิติที่ 1 (L1_2)

จากแผนภาพที่ 4.13 แสดงให้เห็นถึงลักษณะของการกระจายค่า RMSE สำหรับอำนาจจำแนกในมิติที่ 1 จากการทำซ้ำ 20 ครั้ง คล้ายคลึงกันทั้งสองวิธีในแต่ละเงื่อนไข โดยวิธี NOP มีช่วงการกระจายน้อยกว่าวิธี M สำหรับทุกเงื่อนไขที่ศึกษา นอกจากนี้พบว่าเมื่อความสัมพันธ์ระหว่างมิติอยู่ในระดับปานกลางในโครงสร้างที่มีความซับซ้อนมากขึ้นจะมีกระจายของค่า

RMSE น้อยกว่ารูปแบบอื่นๆ ซึ่งวิธี NOP มีการกระจายน้อยกว่าวิธี M แสดงว่าผลที่ได้จากการศึกษาโดยวิธี NOP มีความคงที่ (Stable) มากกว่าวิธี M อย่างไรก็ตามจากแผนภูมิพบว่าค่ากลางของวิธี NOP มีค่าน้อยกว่าวิธี M แสดงว่าโดยเฉลี่ยแล้ววิธี NOP มีความคลาดเคลื่อนของการประมาณค่าการแปลงคะแนนน้อยกว่าวิธี M โดยเฉพาะอย่างยิ่งในเงื่อนไขของมิติความสามารถที่มีความสัมพันธ์กันในระดับสูง

2) ดัชนี RMSE สำหรับค่าอำนาจจำแนกของข้อสอบในมิติที่ 2 (a_2)

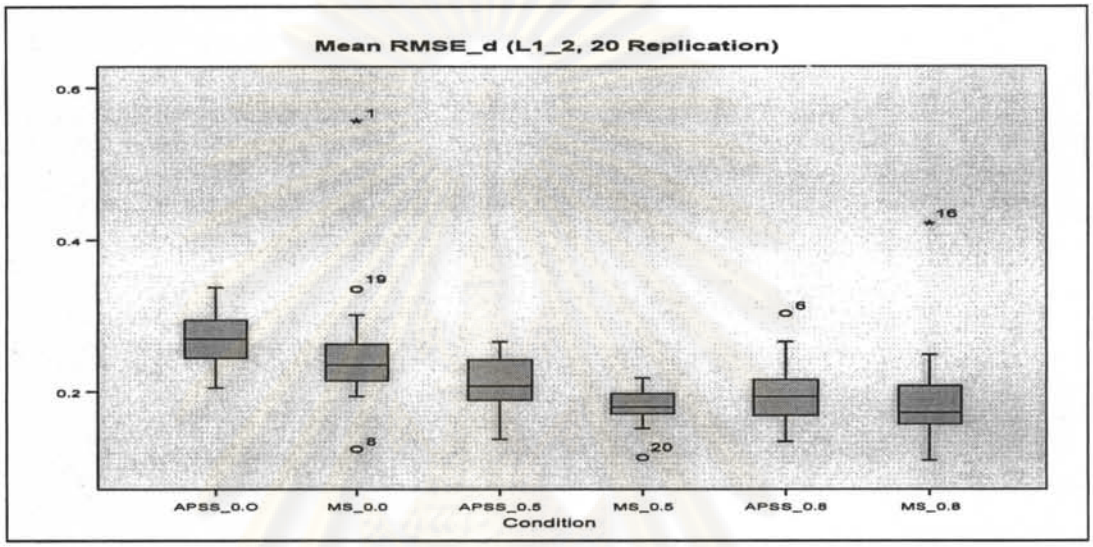


แผนภาพที่ 4.14 แผนภูมิ Box Plot ของข้อมูลชุดวัดซ้ำทั้งหมด 20 ครั้ง สำหรับดัชนี RMSE ของค่าอำนาจจำแนกของข้อสอบในมิติที่ 2 (L1_2)

จากแผนภาพที่ 4.14 โดยส่วนใหญ่มีลักษณะของการกระจายค่า RMSE จากการทำซ้ำ 20 ครั้ง ใกล้เคียงกันระหว่างวิธี M และ วิธี NOP ในแต่ละเงื่อนไขที่ศึกษา ซึ่งส่วนใหญ่วิธี NOP มีช่วงการกระจายน้อยกว่าวิธี M อย่างไรก็ตามพบว่าเมื่อความสัมพันธ์ระหว่างมิติอยู่ในระดับกลางในโครงสร้างที่ซับซ้อนมากขึ้น (MS_0.0) พบว่าวิธี M มีช่วงการกระจายน้อยกว่าวิธี NOP อย่างเห็นได้ชัด แสดงว่าในเงื่อนไขดังกล่าววิธี M มีความคงที่มากกว่า นอกจากนี้พบว่าเมื่อความสัมพันธ์ระหว่างมิติอยู่ในระดับสูงในโครงสร้างที่มีความซับซ้อนมากขึ้น วิธี M จะมีกระจายของค่า RMSE สูงกว่าเงื่อนไขอื่น แสดงว่ามีความคงที่น้อยกว่ารูปแบบอื่น เมื่อพิจารณาค่ากลางพบว่า วิธี NOP มีค่าน้อยกว่าวิธี M ในทุกเงื่อนไขที่ศึกษา แสดงว่าโดยเฉลี่ยแล้ววิธี NOP มีความคลาดเคลื่อนของการประมาณค่าการแปลงคะแนนน้อยกว่าวิธี M โดยเฉพาะอย่างยิ่งในเงื่อนไขของมิติความสามารถที่มีความสัมพันธ์กันสูงในโครงสร้างที่มีความซับซ้อนมากขึ้น

โดยสรุปสำหรับค่าอำนาจจำแนกทั้ง 2 มิติ พบว่าลักษณะการกระจายของดัชนี RMSE ไม่แตกต่างกันมากนัก โดยเมื่อความสัมพันธ์ระหว่างมิติความสามารถสูงขึ้นมีแนวโน้มที่จะมีค่าดัชนี RMSE รวมถึงค่ากลางเพิ่มขึ้นโดยเฉพาะวิธี M และผลการพิจารณาโดยภาพรวม พบว่าวิธี NOP มีความคงที่ในการประมาณค่าที่ดีกว่าวิธี M เมื่อพิจารณาค่ากลางพบว่า วิธี NOP มีค่าน้อยกว่าวิธี M ในทุกเงื่อนไขที่ศึกษา

3) ดัชนี RMSE สำหรับค่าความยากข้อสอบ (d)



แผนภาพที่ 4.15 แผนภูมิ Box Plot ของข้อมูลชุดวัดซ้ำทั้งหมด 20 ครั้ง สำหรับดัชนี RMSE ของค่าความยากของข้อสอบ (L1_2)

เมื่อพิจารณาดัชนี RMSE สำหรับค่าพารามิเตอร์ d พบว่า โครงสร้างมิติความสามารถที่ไม่ซับซ้อนมีช่วงของค่า RMSE สูงกว่าโครงสร้างที่ซับซ้อน โดยเมื่อความสัมพันธ์ระหว่างมิติเพิ่มสูงขึ้นจะมีค่าค่า RMSE สูงขึ้น นั่นคือมีความคลาดเคลื่อนของการประมาณค่าน้อยนั่นเอง โดยเมื่อความสัมพันธ์ในระดับปานกลางที่มีโครงสร้างไม่ซับซ้อนพบว่ามีการกระจายของคะแนนน้อยกว่าเงื่อนไขอื่นอย่างเห็นได้ชัด แสดงว่ามีความคงที่ของดัชนี RMSE มากกว่า

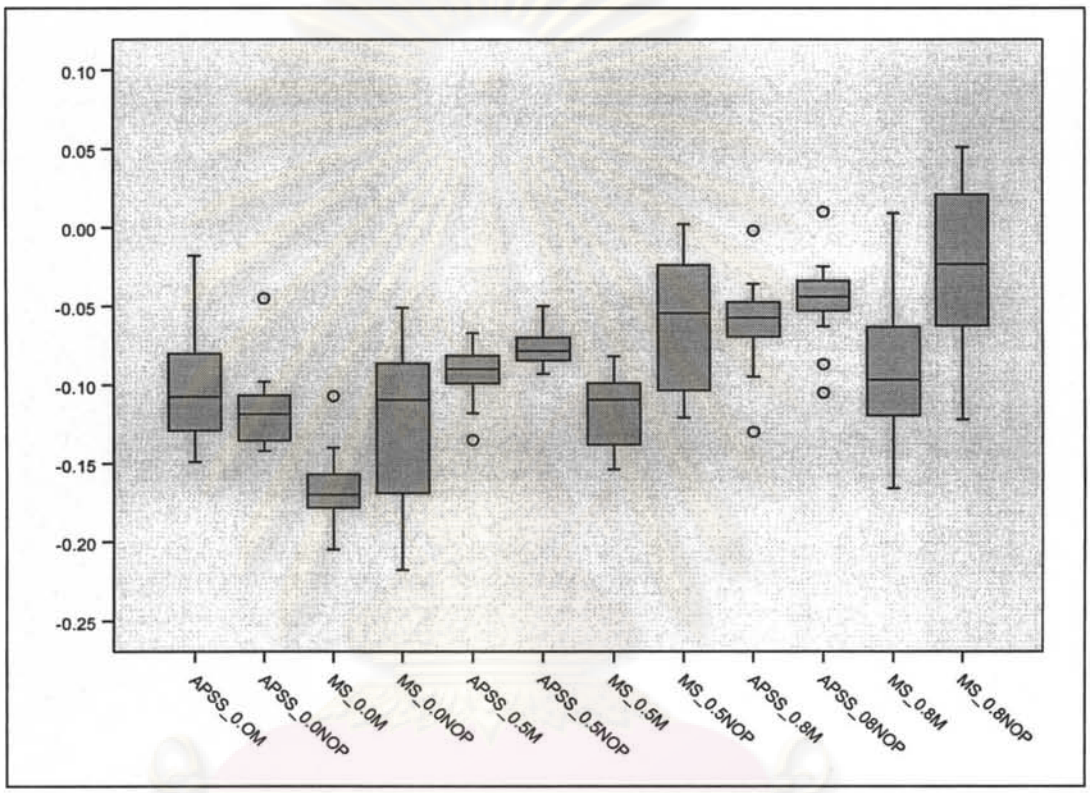
2.1.2 ผลการวิเคราะห์โดยใช้ดัชนี BIAS เป็นตัวแปรตาม พบว่า เกือบทุกเงื่อนไขมีผลทำให้ค่าเฉลี่ยดัชนี BIAS ของวิธี M แตกต่างจากวิธี NOP อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยทั้งสองวิธีมีการประมาณค่าที่ต่ำกว่าความเป็นจริง แต่วิธี M จะมีการประมาณค่าที่ต่ำกว่าความเป็นจริงมากกว่าวิธี NOP อย่างเห็นได้ชัด มีเพียงเงื่อนไขโครงสร้างที่ไม่ซับซ้อนของเงื่อนไขความสัมพันธ์ระหว่างมิติสูง หรือไม่มีความสัมพันธ์ระหว่างมิติ (APSS_0.0 และ APSS_0.8) ที่มีค่าเฉลี่ยดัชนี BIAS ของวิธี M แตกต่างจากวิธี NOP อย่างไม่มีนัยสำคัญทางสถิติ

เมื่อพิจารณาโดยภาพรวม พบว่า โครงสร้างที่ซับซ้อนมากขึ้น (MS) มีการประมาณค่าต่ำกว่าความเป็นจริงสูงกว่าโครงสร้างที่ไม่ซับซ้อน (APSS) ในทุกระดับความสัมพันธ์สำหรับวิธี M ซึ่งให้ผลที่มีลักษณะใกล้เคียงกับวิธี NOP แต่วิธี NOP ให้ผลที่แตกต่างจากวิธี M คือ

เมื่อความสัมพันธ์ระหว่างมิติอยู่ในระดับสูง โครงสร้างที่ไม่ซับซ้อนจะมีการประมาณค่าที่ต่ำกว่าความเป็นจริงมากกว่าโครงสร้างที่ซับซ้อน

จากที่กล่าวมาสามารถแสดงแผนภาพ Box Plot เพื่อพิจารณาลักษณะการกระจายของค่าความคลาดเคลื่อน BIAS ของค่าพารามิเตอร์การแปลงคะแนนในมิติที่ 1 และ 2 ดังต่อไปนี้

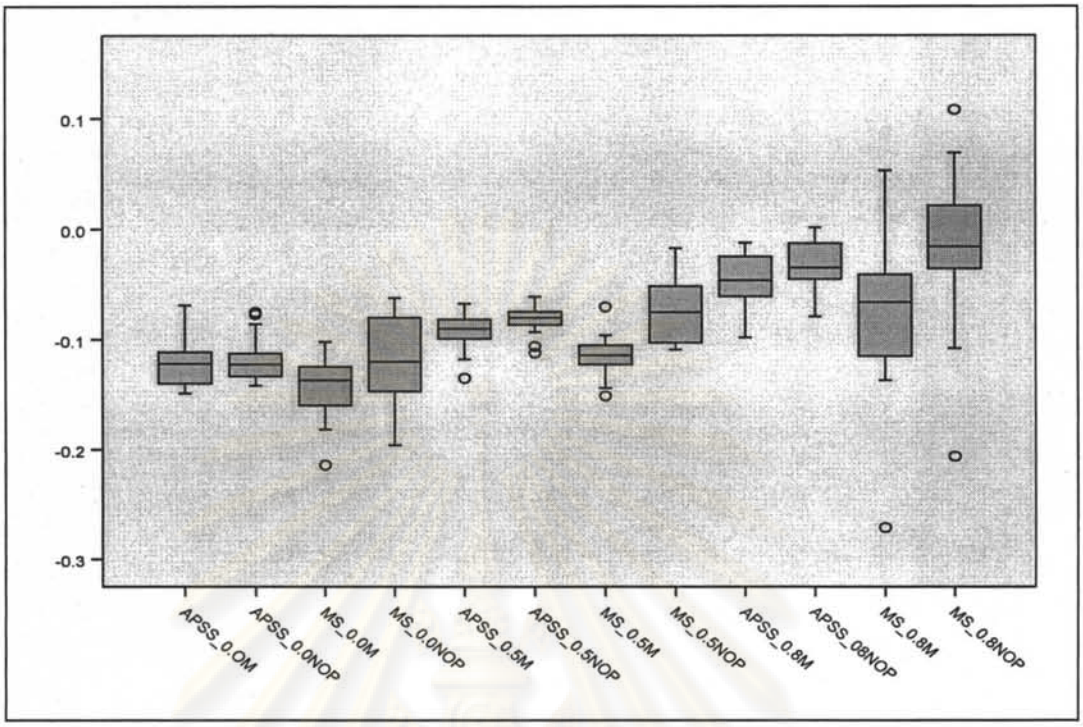
1) ดัชนี BIAS สำหรับค่าอำนาจจำแนกของข้อสอบในมิติที่ 1 (a_1)



แผนภาพที่ 4.16 แผนภูมิ Box Plot ของข้อมูลชุดวัดซ้ำทั้งหมด 20 ครั้ง สำหรับดัชนี BIAS ของค่าอำนาจจำแนกของข้อสอบในมิติที่ 1 (L1_2)

จากแผนภาพที่ 4.16 แสดงให้เห็นถึงลักษณะของการกระจายค่า BIAS สำหรับอำนาจจำแนกในมิติที่ 1 จากการทำซ้ำ 20 ครั้ง เมื่อพิจารณารายเงื่อนไข พบว่ามีลักษณะของการกระจายค่า BIAS แตกต่างกันอย่างชัดเจนระหว่างสองวิธี บางเงื่อนไขวิธี M มีการกระจายน้อยกว่า NOP เช่น MS_0.0, MS_0.5 และ MS_0.8 ในขณะที่โครงสร้างไม่ซับซ้อน วิธี NOP มีการกระจายน้อยกว่า M โดยเฉพาะอย่างยิ่งในเงื่อนไขของความสัมพันธ์ระหว่างมิติระดับปานกลางจะมีค่าการกระจายของดัชนี BIAS น้อยกว่าเงื่อนไขอื่น เมื่อพิจารณาโดยภาพรวมพบว่า การกระจายของทั้งสองวิธีจะประมาณค่าต่ำกว่าความเป็นจริงทั้งสองวิธี แต่วิธี M จะมีความคลาดเลื่อนสูงกว่าวิธี NOP อย่างชัดเจน อย่างไรก็ตามเมื่อความสัมพันธ์ระหว่างมิติอยู่ในระดับสูงสำหรับวิธี NOP จะมีการประมาณค่าใกล้เคียงกับความเป็นจริงมากกว่าเงื่อนไขอื่น

2) ดัชนี BIAS สำหรับค่าอำนาจจำแนกของข้อสอบในมิติที่ 2 (a2)

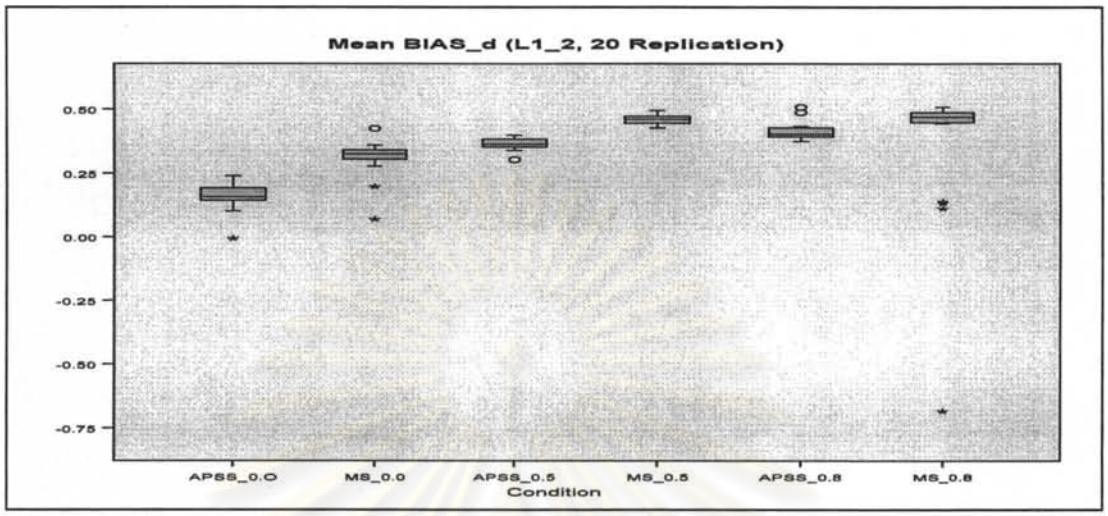


แผนภาพที่ 4.17 แผนภูมิ Box Plot ของข้อมูลชุดวัดซ้ำทั้งหมด 20 ครั้ง สำหรับดัชนี BIAS ของค่าอำนาจจำแนกของข้อสอบในมิติที่ 2 (L1_2)

จากแผนภาพที่ 4.17 เมื่อพิจารณารายเงื่อนไข พบว่าเมื่อความสัมพันธ์ระหว่างมิติเพิ่มขึ้นจะมีลักษณะของการกระจายค่า BIAS แตกต่างกันระหว่างสองวิธี ซึ่งโดยส่วนใหญ่วิธี NOP มีการกระจายน้อยกว่าวิธี M มีเพียง 2 เงื่อนไข ที่วิธี M มีการกระจายน้อยกว่า นั่นคือ MS_0.0 และ MS_0.5 แสดงว่าในเงื่อนไขดังกล่าววิธี M มีความคงที่ของดัชนี BIAS มากกว่าวิธี NOP เมื่อพิจารณาโดยภาพรวมพบว่า การกระจายของทั้งสองวิธีจะประมาณค่าต่ำกว่าความเป็นจริง แต่วิธี M จะมีความคลาดเคลื่อนสูงกว่าวิธี NOP อย่างชัดเจน อย่างไรก็ตามเมื่อความสัมพันธ์ระหว่างมิติอยู่ในระดับสูงสำหรับวิธี NOP จะมีการประมาณค่าใกล้เคียงกับความเป็นจริงมากกว่าเงื่อนไขอื่น โดยเมื่อความสัมพันธ์ระหว่างมิติความสามารถสูง วิธี NOP จะประมาณค่าได้ใกล้เคียงกับค่าที่แท้จริงมากกว่าเงื่อนไขอื่นๆ และมีแนวโน้มของการกระจายที่มีการประมาณค่าสูงกว่าความเป็นจริง

โดยสรุปสำหรับค่าอำนาจจำแนกทั้ง 2 มิติ พบว่าลักษณะการกระจายของดัชนี BIAS มีลักษณะที่ไม่แตกต่างกันมากนัก โดยเมื่อความสัมพันธ์ระหว่างมิติความสามารถสูงขึ้นมีแนวโน้มที่จะมีค่าดัชนี BIAS ลดลง ซึ่งโดยส่วนใหญ่วิธี M มีการประมาณค่าต่ำกว่าความเป็นจริงมากกว่าวิธี NOP ในขณะที่เมื่อความสัมพันธ์ระหว่างมิติสูงขึ้น วิธี NOP มีแนวโน้มในการประมาณค่ามากกว่าความเป็นจริง

3) ดัชนี BIAS สำหรับค่าความยากของข้อสอบ (d)



แผนภาพที่ 4.18 แผนภูมิ Box Plot ของข้อมูลชุดวัดซ้ำทั้งหมด 20 ครั้ง สำหรับดัชนี RMSE ของค่าความยากของข้อสอบ (L1_2)

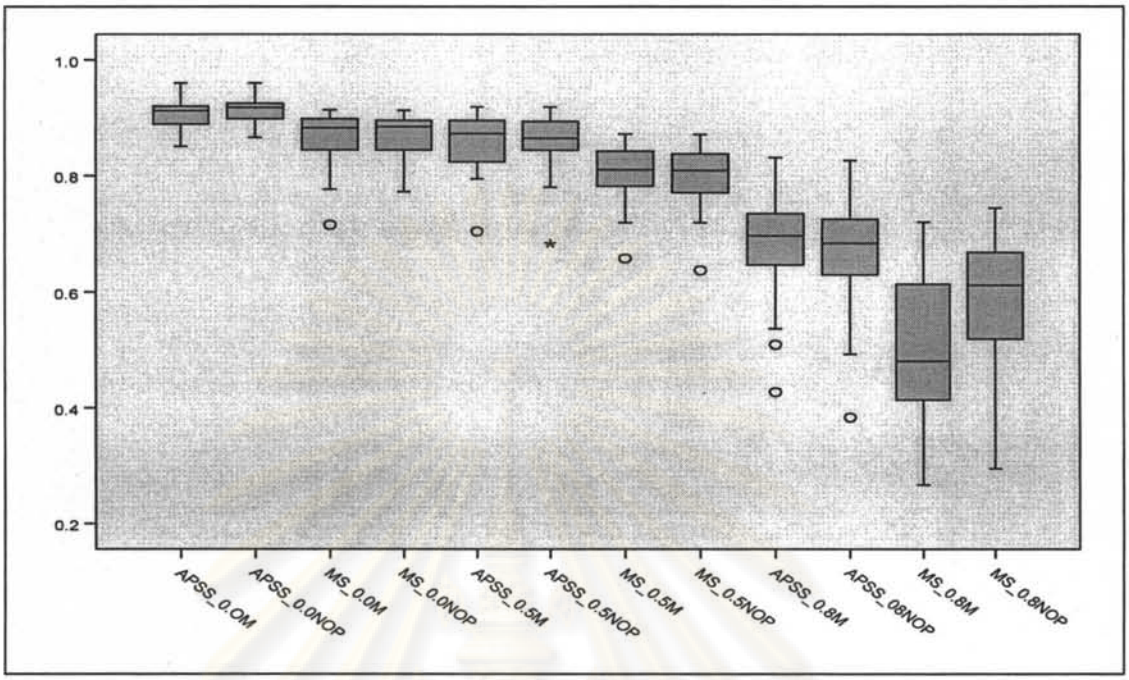
จากแผนภาพที่ 4.18 สำหรับค่าดัชนี BIAS เมื่อความสัมพันธ์ระหว่างมิติเพิ่มขึ้นมีผลทำให้ประมาณค่าสูงกว่าความเป็นจริงเพิ่มสูงขึ้น โดยโครงสร้างมิติความสามารถที่ไม่ซับซ้อนมีการประมาณค่าใกล้เคียงกับค่าที่แท้จริงมากกว่าโครงสร้างที่ซับซ้อน นอกจากนี้พบว่าเมื่อความสัมพันธ์ในระดับปานกลางที่มีโครงสร้างไม่ซับซ้อนพบว่ามีการกระจายของคะแนนน้อยกว่าเงื่อนไขอื่นอย่างเห็นได้ชัด แสดงว่ามีความคงที่ของดัชนี BIAS มากกว่า

2.1.3 ผลการวิเคราะห์โดยใช้ค่า CORR เป็นตัวแปรตาม พบว่ามีเพียงเงื่อนไข MS_0.8 เท่านั้นที่มีผลทำให้ค่าสัมประสิทธิ์สหสัมพันธ์ของวิธี M แตกต่างจากวิธี NOP อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ทั้งค่าพารามิเตอร์ a_1 (t=-2.348) และ a_2 (t=-2.23) สำหรับเงื่อนไขอื่นๆ พบว่ามีค่าเฉลี่ยของ CORR แตกต่างกันแต่ไม่มีนัยสำคัญทางสถิติ

อย่างไรก็ตามแม้ผลการวิเคราะห์พบว่าทั้ง 2 วิธีให้ผลไม่แตกต่างกันในแต่ละเงื่อนไข แต่ข้อค้นพบในการศึกษาจากค่าสัมประสิทธิ์สหสัมพันธ์ของค่าพารามิเตอร์ a_1 และ a_2 พบว่า โครงสร้างที่ไม่ซับซ้อนมีค่าสัมประสิทธิ์สหสัมพันธ์สูงกว่าในทุกระดับความสัมพันธ์ โดยเมื่อความสัมพันธ์ระหว่างมิติสูงขึ้นจะมีค่า CORR ลดน้อยลง

จากที่กล่าวมาสามารถแสดงแผนภาพ Box Plot เพื่อพิจารณาลักษณะการกระจายของค่าความคลาดเคลื่อน BIAS ของค่าพารามิเตอร์การแปลงคะแนนในมิติที่ 1 และ 2 ดังต่อไปนี้

1) ค่า CORR สำหรับค่าอำนาจจำแนกของข้อสอบในมิติที่ 1 (a_1)

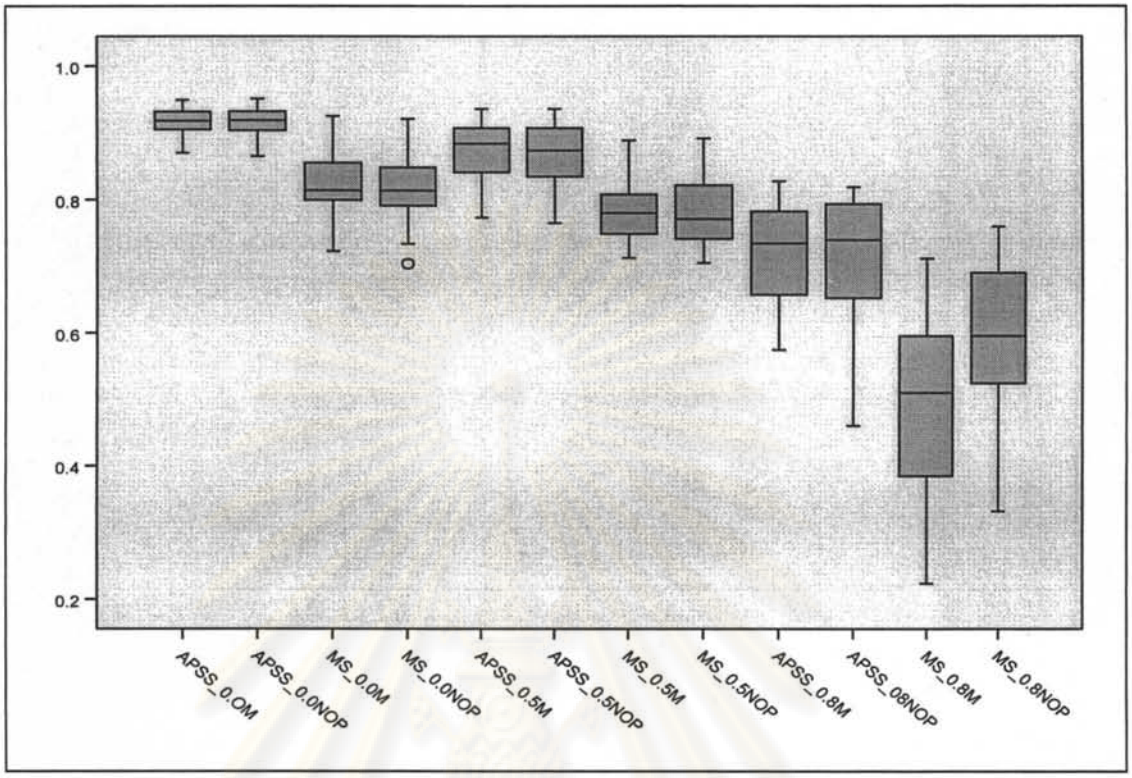


แผนภาพที่ 4.19 แผนภูมิ Box Plot ของข้อมูลชุดวัดซ้ำทั้งหมด 20 ครั้ง สำหรับ CORR ของค่าอำนาจจำแนกของข้อสอบในมิติที่ 1 (L1_2)

จากแผนภาพที่ 4.19 แสดงให้เห็นถึงลักษณะของการกระจายค่า CORR สำหรับอำนาจจำแนกในมิติที่ 1 จากการทำซ้ำ 20 ครั้ง คล้ายคลึงกันทั้งสองวิธีในแต่ละเงื่อนไข โดยส่วนใหญ่วิธี NOP มีช่วงการกระจายน้อยกว่าวิธี M มีเพียงเงื่อนไขที่ความสัมพันธ์ระหว่างมิติในระดับปานกลางในโครงสร้างที่ซับซ้อน (MS_0.5) พบว่า วิธี NOP มีช่วงการกระจายน้อยกว่าวิธี M ถ้าพิจารณาโดยภาพรวมพบว่าเมื่อความสัมพันธ์ระหว่างมิติเพิ่มสูงขึ้นมีแนวโน้มว่าค่า CORR ลดลงอย่างเห็นโดยชัด รวมถึงมีการกระจายของค่าสัมประสิทธิ์สหสัมพันธ์เพิ่มขึ้น เมื่อพิจารณาค่ากลางพบว่าทั้งสองวิธีมีค่าสัมประสิทธิ์สหสัมพันธ์ใกล้เคียงกันแต่เมื่อความสัมพันธ์ระหว่างมิติอยู่ในระดับสูงที่มีโครงสร้างซับซ้อนมากขึ้น จะมีค่าอยู่ในระดับที่ไม่สูงนัก โดยวิธี NOP มีค่าสูงกว่าวิธี M อย่างเห็นได้ชัด

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

2) ค่า CORR สำหรับค่าอำนาจจำแนกของข้อสอบในมิติที่ 2 (a2)

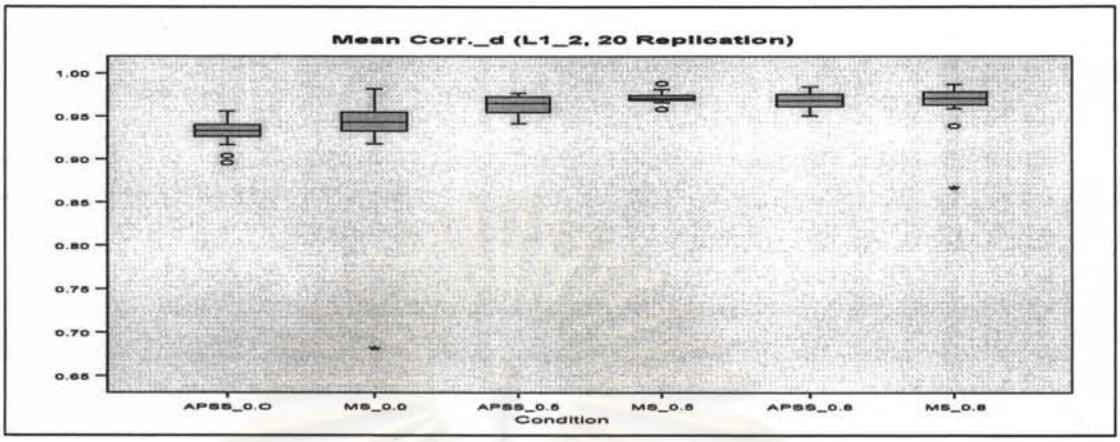


แผนภาพที่ 4.20 แผนภูมิ Box Plot ของข้อมูลชุดวัดซ้ำทั้งหมด 20 ครั้ง สำหรับค่า CORR ของค่าอำนาจจำแนกของข้อสอบในมิติที่ 2 (L1_2)

จากแผนภาพที่ 4.16 เมื่อพิจารณารายเงื่อนไข พบว่า ในแต่ละเงื่อนไขของทั้งสองวิธีมีความคล้ายคลึงกัน ยกเว้นเงื่อนไข MS_0.8 วิธีทั้งสองมีการกระจายของช่วงคะแนนที่แตกต่างกัน ซึ่งวิธี M มีการกระจายของช่วงคะแนนมากกว่าวิธี NOP อย่างเห็นได้ชัด สำหรับการพิจารณาโดยภาพรวมพบว่าเมื่อความสัมพันธ์ระหว่างมิติเพิ่มขึ้นมีแนวโน้มว่าค่า CORR จะลดลงอย่างเห็นได้ชัด รวมถึงมีการกระจายของค่า CORR เพิ่มขึ้น เมื่อพิจารณาค่ากลางพบว่าทั้งสองวิธีมีค่า CORRที่ใกล้เคียงกันแต่เมื่อระดับความสัมพันธ์ระหว่างมิติในระดับสูงสำหรับโครงสร้างที่ซับซ้อนมากขึ้น จะมีค่าอยู่ในระดับที่ไม่สูงนัก โดยวิธี NOP มีค่าสูงกว่าวิธี M อย่างเห็นได้ชัด

โดยสรุปสำหรับค่าอำนาจจำแนกทั้ง 2 มิติ พบว่าลักษณะการกระจายของค่า CORR มีลักษณะที่ไม่แตกต่างกันมากนัก โดยเมื่อความสัมพันธ์ระหว่างมิติความสามารถสูงขึ้นมีแนวโน้มที่จะมีค่า CORR ลดลงอย่างเห็นได้ชัด โดยวิธี NOP มีค่า CORR สูงกว่าวิธี M

3) ค่า CORR สำหรับค่าความยากของข้อสอบ (d)



แผนภาพที่ 4.21 แผนภูมิ Box Plot ของข้อมูลชุดวัดซ้ำทั้งหมด 20 ครั้ง สำหรับดัชนี RMSE BIAS และ CORR สำหรับค่าพารามิเตอร์ d รายเงื่อนไขที่ศึกษา (L1_2)

จากแผนภาพ 4.17 เมื่อพิจารณาค่า CORR พบว่าเมื่อความสัมพันธ์ระหว่างมิติความสามารถเพิ่มขึ้นมีผลทำให้ค่า CORR เพิ่มสูงขึ้น โดยทุกเงื่อนไขมีค่าใกล้ 1 โดยเงื่อนไขความสัมพันธ์ระหว่างมิติอยู่ในระดับกลางที่มีโครงสร้างที่ไม่ซับซ้อนมีความคงที่ของค่า CORR มากกว่าเงื่อนไขอื่น

โดยสรุปพบว่าส่วนใหญ่วิธี M มีความคลาดเคลื่อนของการประมาณค่าและมีการประมาณค่าที่ต่ำกว่าความเป็นจริงมากกว่าวิธี NOP โดยโครงสร้างที่ซับซ้อนมีความคลาดเคลื่อนสูงกว่าและมีการประมาณค่าที่ต่ำกว่าค่าที่แท้จริงมากกว่าโครงสร้างที่ไม่ซับซ้อน ในขณะที่ความสัมพันธ์เกือบทุกเงื่อนไขมีค่าเข้าใกล้ 1

2.2 ผลการเปรียบเทียบค่าเฉลี่ยของดัชนี RMSE BIAS และ CORR จากระดับที่ 2 ไปยังระดับที่ 3 รายสถานการณ์

2.2.1 ผลการเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนนระหว่างวิธี M กับวิธี NOP ของแต่ละโครงสร้างมิติความสามารถ (APSS และ MS) และแต่ละระดับความสัมพันธ์มิติความสามารถ (r=0.0, r=0.5 และ r=0.8) โดยใช้สถิติทดสอบ One-way MANOVA ได้ผลดังตารางที่ 4.15

จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.15 ผลการทดสอบความแปรปรวนแบบพหุจำแนกทางเดียว (One-way MANOVA) เพื่อเปรียบเทียบค่าเฉลี่ยของดัชนี Ln RMSE BIAS และค่า CORR ระหว่างวิธี M และวิธี NOP ในแต่ละเงื่อนไขที่ศึกษา จากระดับที่ 2 ไปยังระดับที่ 3 (L2_3)

เงื่อนไขการจำลองข้อมูล		วิธีการ	Ln RMSE				BIAS				CORR			
			a1		a2		a1		a2		a1		a2	
			M (SD)	F	M (SD)	F	M (SD)	F	M (SD)	F	M (SD)	F	M (SD)	F
DS	APSS	M	-2.089 (0.241)	0.552	-2.076 (0.185)	0.822	-0.109 (0.044)	1.040	-0.114 (0.045)	0.288	0.859 (0.106)	0.064	0.850 (0.100)	0.004
		NOP	-2.121 (0.234)		-2.106 (0.177)		-0.101 (0.045)		-0.109 (0.050)		0.854 (0.111)		0.849 (0.100)	
	MS	M	-2.001 (0.282)	1.831	-2.024 (0.238)	2.712	-0.152 (0.052)	**4.905***	-0.138 (0.057)	2.720	0.765 (0.177)	0.267	0.774 (0.132)	0.144
		NOP	-2.065 (0.234)		-2.092 (0.213)		-0.127 (0.070)		-0.119 (0.069)		0.781 (0.159)		0.783 (0.127)	
r	r=0.0	M	-2.096 (0.206)	0.054	-2.109 (0.139)	0.133	-0.186 (0.035)	0.340	-0.189 (0.022)	0.073	0.916 (0.040)	0.016	0.910 (0.037)	0.077
		NOP	-2.107 (0.202)		-2.120 (0.133)		-0.182 (0.033)		-0.191 (0.021)		0.915 (0.044)		0.912 (0.031)	
	r=0.5	M	-2.132 (0.238)	0.267	-2.101 (0.196)	0.928	-0.106 (0.019)	**22.172****	-0.106 (0.018)	**24.844****	0.866 (0.073)	0.014	0.837 (0.078)	0.012
		NOP	-2.159 (0.233)		-2.143 (0.199)		-0.077 (0.035)		-0.084 (0.023)		0.868 (0.076)		0.839 (0.077)	
	r=0.8	M	-1.906 (0.292)	3.100	-1.940 (0.252)	3.016	-0.098 (0.045)	2.841	-0.083 (0.035)	3.541	0.655 (0.156)	0.221	0.690 (0.113)	0.098
		NOP	-2.013 (0.248)		-2.033 (0.229)		-0.082 (0.039)		-0.069 (0.032)		0.670 (0.137)		0.698 (0.109)	

*p< .05, **p<.01

ตารางที่ 4.16 ผลการทดสอบรายคู่ที่มีนัยสำคัญทางสถิติที่ระดับ .05

เงื่อนไขที่ศึกษา		RMSE		BIAS		CORR	
		a_1	a_2	a_1	a_2	a_1	a_2
โครงสร้างมิติความสามารถ (DS)	APSS	X	X	X	X	X	X
	MS	X	X	**M>NOP**	X	X	X
ระดับความสัมพันธ์ระหว่างมิติความสามารถ (r)	r=0.0	X	X	X	X	X	X
	r=0.5	X	X	**M>NOP**	**M>NOP**	X	X
	r=0.8	X	X	X	X	X	X

X ไม่พบว่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติ

จากตารางที่ 4.15 และ 4.16 สามารถสรุปได้ว่าการเชื่อมโยงคะแนนสำหรับชุดของข้อสอบที่มีโครงสร้างไม่ซับซ้อน (APSS) พบว่าทั้ง 2 วิธี มีคุณภาพของการเชื่อมโยงคะแนนไม่แตกต่างกัน แต่เมื่อโครงสร้างมีความซับซ้อนมากขึ้น (MS) พบว่าวิธี M มีค่าดัชนี BIAS มากกว่าวิธี NOP สะท้อนให้เห็นว่าวิธี NOP มีความคงที่ของการเชื่อมโยงคะแนนมากกว่าวิธี M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 (t=4.905)

เมื่อพิจารณาระดับความสัมพันธ์ระหว่างมิติความสามารถ พบว่าความสัมพันธ์ระหว่างมิติความสามารถอยู่ในระดับกลางวิธี NOP จะมีค่า BIAS ต่ำกว่าวิธี M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ทั้งในมิติที่ 1 (t=22.172) และมิติที่ 2 (t=24.844) สะท้อนให้เห็นว่าวิธี NOP มีความถูกต้องในการประมาณค่ามากกว่าวิธี M อย่างไรก็ตามจากผลการวิจัยพบว่าระดับความสัมพันธ์ระหว่างมิติความสามารถไม่ส่งผลกระทบต่อค่า RMSE ซึ่งสะท้อนถึงความคงที่ของการเชื่อมโยงคะแนน

ข้อสังเกตที่ได้จากผลการวิจัยครั้งนี้พบว่า ค่า CORR ทั้ง 2 วิธี ไม่ต่างกัน ไม่ว่าจะศึกษาจากโครงสร้างมิติความสามารถลักษณะใด และ ระดับความสัมพันธ์ระหว่างมิติความสามารถระดับใด แสดงว่าทั้งสองวิธีมีความสัมพันธ์ระหว่างค่าที่แปลงได้จากแบบสอบปรับเทียบคะแนนกับคะแนนเกณฑ์ที่ได้จากแบบสอบฐานไม่ต่างกัน

2.2.2 ผลการเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนนระหว่างวิธี M กับวิธี NOP ภายใต้ 6 เงื่อนไข คือ APSS_0.0, MS_0.0, APSS_0.0, MS_0.5, APSS_0.8 และ MS_0.8 โดยใช้สถิติทดสอบ t ปรากฏผลดังตารางที่ 4.17

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.17 การเปรียบเทียบค่าเฉลี่ยของดัชนี RMSE BIAS และค่า CORR ระหว่างวิธี M และวิธี NOP จากระดับที่ 2 ไปยังระดับที่ 3 (L2_3)

เงื่อนไขการจำลองข้อมูล	วิธีการ	RMSE				BIAS				Correlation Coefficient			
		a_1		a_2		a_1		a_2		a_1		a_2	
		M (SD)	t	M (SD)	t	M (SD)	t	M (SD)	t	M (SD)	t	M (SD)	t
Corr. 0.0 APSS	M	0.132	0.196	0.124	0.219	-0.162	-0.869	-0.172	0.556	0.933	0.101	0.932	-0.171
		(0.025)		(0.016)		(0.025)		(0.015)		(0.019)		(0.017)	
(APSS_0.0)	NOP	0.130		0.123		-0.156		-0.174	0.556	0.933		0.933	
		(0.025)		(0.016)		(0.019)		(0.011)		(0.019)		(0.017)	
Corr. 0.0 MS	M	0.120	0.142	0.121	0.328	-0.211	-0.375	-0.207	0.065	0.899	0.109	0.888	-0.311
		(0.028)		(0.018)		(0.025)		(0.011)		(0.048)		(0.039)	
(MS_0.0)	NOP	0.118		0.119		0.328		-0.207		0.897		0.891	
		(0.027)		(0.016)		(0.021)		(0.014)		(0.053)		(0.027)	
Corr. 0.5 APSS	M	0.110	0.233	0.119	0.257	-0.098	-1.709	-0.098	-1.201	0.904	0.116	0.887	0.158
		(0.031)		(0.026)		(0.013)		(0.013)		(0.062)		(0.045)	
(APSS_0.5)	NOP	0.108		0.117		0.257		-0.094		0.901		0.884	
		(0.031)		(0.026)		(0.012)		(0.008)		(0.064)		(0.050)	
Corr. 0.5 MS	M	0.134	0.819	0.130	1.447	-0.115	**-4.840****	-0.115	**-5.567****	0.827	-0.292	0.787	-0.269
		(0.020)		(0.017)		(0.021)		(0.018)		(0.063)		(0.073)	
(MS_0.5)	NOP	0.129		0.122		-0.062		-0.073		0.834		0.794	
		(0.015)		(0.018)		(0.043)		(0.028)		(0.073)		(0.074)	
Corr. 0.8 APSS	M	0.141	1.168	0.139	1.556	-0.067	**-2.292***	-0.072	**-2.484***	0.739	0.438	0.733	0.085
		(0.026)		(0.018)		(0.019)		(0.016)		(0.086)		(0.075)	
(APSS_0.8)	NOP	0.132		0.131		-0.055		-0.060		0.727		0.731	
		(0.022)		(0.017)		(0.014)		(0.015)		(0.089)		(0.070)	
Corr. 0.8 MS	M	0.167	**2.177****	0.156	1.788	-0.130	-1.593	-0.094	-1.180	0.570	-0.849	0.648	-0.423
		(0.041)		(0.037)		(0.041)		(0.045)		(0.166)		(0.129)	
(MS_0.8)	NOP	0.142		0.137		1.788		-0.077		0.613		0.665	
		(0.031)		(0.030)		(0.038)		(0.041)		(0.155)		(0.131)	

*p<.0.5, **p<.01

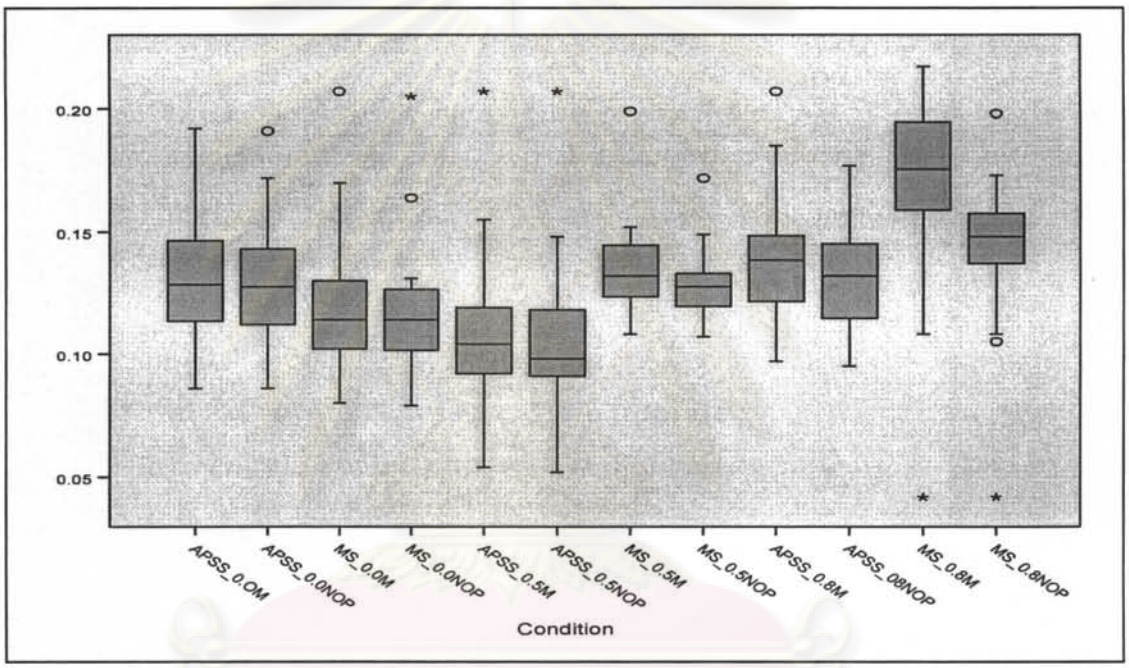
จากตารางที่ 4.17 ผลการเปรียบเทียบค่าเฉลี่ยของดัชนีบ่งชี้คุณภาพแต่ละชนิดในแต่ละระดับของตัวแปรเงื่อนไขของการจำลองข้อมูล ได้ผลดังต่อไปนี้

2.2.1) ผลการวิเคราะห์โดยใช้ดัชนี Ln RMSE เป็นตัวแปรตาม พบว่ามีลักษณะใกล้เคียงกับการเชื่อมโยงคะแนนในระดับที่ 1 ไปยังระดับที่ 2 นั่นคือ มีเพียงเงื่อนไข MS_0.8 เท่านั้นที่มีผลทำให้ค่าดัชนี RMSE ของวิธี M แตกต่างจากวิธี NOP แต่จะแตกต่างอย่างมีนัยสำคัญทางสถิติเฉพาะค่าอำนาจจำแนกในมิติที่ 1 (t=2.177) โดยวิธี NOP (RMSE_a_1=0.142, RMSE_a_2=0.137) มีความคงที่ของการประมาณค่าสูงกว่าวิธี M (RMSE_a_1=0.167, RMSE_a_2=0.156) อย่างเห็นได้ชัด

สำหรับเงื่อนไขอื่นๆ พบว่ามีค่าเฉลี่ยของดัชนี RMSE แตกต่างกันแต่ไม่มีนัยสำคัญทางสถิติ อย่างไรก็ตามแม้ผลการวิเคราะห์พบว่าทั้ง 2 วิธีให้ผลไม่แตกต่างกันในแต่ละเงื่อนไข แต่ข้อค้นพบในการศึกษาจากค่าเฉลี่ยของดัชนี RMSE ของค่าพารามิเตอร์ a_1 และ a_2 พบว่า เงื่อนไข APSS_0.5 มีความคลาดเคลื่อนน้อยกว่าความสัมพันธ์ระดับอื่น

อย่างไรก็ตามเมื่อความสัมพันธ์ระหว่างมิติความสามารถในระดับสูง (r=0.8) พบว่ามีความคลาดเคลื่อนของการประมาณค่ามากกว่าระดับความสัมพันธ์อื่นๆ โดยเฉพาะวิธี M จากที่กล่าวมาสามารถแสดงแผนภาพ Box Plot เพื่อพิจารณาลักษณะการกระจายของดัชนี RMSE ของค่าพารามิเตอร์การแปลงคะแนนในมิติที่ 1 และ 2 ดังต่อไปนี้

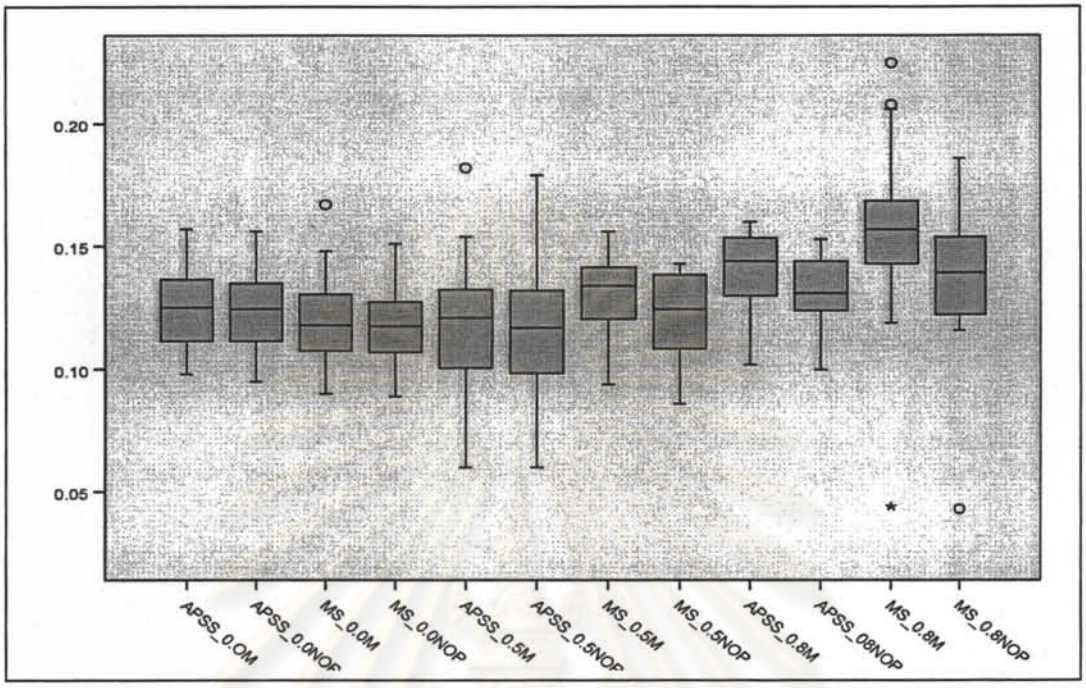
1) ดัชนี RMSE สำหรับค่าอำนาจจำแนกของข้อสอบในมิติที่ 1 (a_1)



แผนภาพที่ 4.22 แผนภูมิ Box Plot ของข้อมูลชุดวัดซ้ำทั้งหมด 20 ครั้ง สำหรับดัชนี RMSE ของค่าอำนาจจำแนกของข้อสอบในมิติที่ 1 (L2_3)

จากแผนภาพที่ 4.22 แสดงให้เห็นถึงลักษณะของการกระจายค่า RMSE สำหรับอำนาจจำแนกในมิติที่ 1 จากการทำซ้ำ 20 ครั้ง พบว่า วิธี NOP มีช่วงการกระจายน้อยกว่าวิธี M โดยโครงสร้างที่ไม่ซับซ้อนมีการกระจายของคะแนนมากกว่าโครงสร้างที่ซับซ้อนมากขึ้น นอกจากนี้พบว่าเมื่อความสัมพันธ์ระหว่างมิติอยู่ในระดับปานกลางในโครงสร้างที่มีความซับซ้อนมากขึ้นจะมีกระจายของค่า RMSE น้อยกว่ารูปแบบอื่นๆ ซึ่งวิธี NOP มีการกระจายน้อยกว่าวิธี M แสดงว่าผลที่ได้จากการศึกษาโดยวิธี NOP มีความคงที่ (Stable) มากกว่าวิธี M อย่างไรก็ตามจากแผนภูมิพบว่าค่ากลางของวิธี NOP มีค่าน้อยกว่าวิธี M แสดงว่าโดยเฉลี่ยแล้ววิธี NOP มีความคลาดเคลื่อนของการประมาณค่าการแปลงคะแนนน้อยกว่าวิธี M โดยเฉพาะอย่างยิ่งในเงื่อนของมิติที่มีความสัมพันธ์กันในระดับสูง ซึ่งให้ผลที่ใกล้เคียงกับการแปลงคะแนนจากระดับที่ 1 ไปยังระดับที่ 2

2) ดัชนี RMSE สำหรับค่าอำนาจจำแนกของข้อสอบในมิติที่ 2 (a_2)

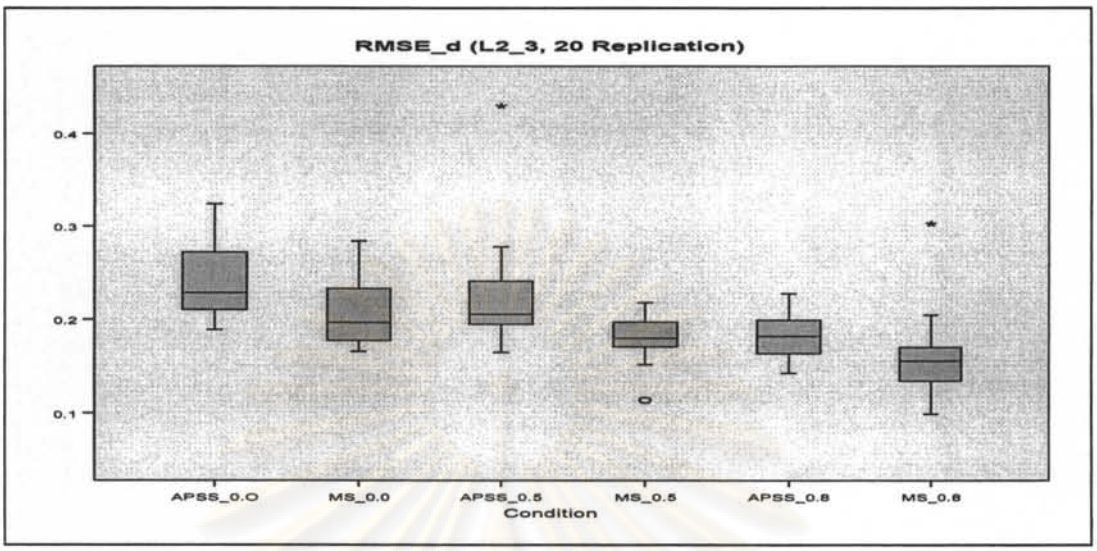


แผนภาพที่ 4.23 แผนภูมิ Box Plot ของข้อมูลชุดวัดซ้ำทั้งหมด 20 ครั้ง สำหรับดัชนี RMSE ของค่าอำนาจจำแนกของข้อสอบในมิติที่ 2 (L2_3)

จากแผนภาพที่ 4.23 โดยส่วนใหญ่มีลักษณะของการกระจายค่า RMSE สำหรับอำนาจจำแนกในมิติที่ 2 จากการทำซ้ำ 20 ครั้ง ใกล้เคียงกันระหว่างวิธี M และ วิธี NOP ในแต่ละเงื่อนไขที่ศึกษา สำหรับความสัมพันธ์ระหว่างมิติความสามารถในระดับปานกลางหรือหรือไม่มีความสัมพันธ์ระหว่างมิติ แต่เมื่อมีความความสัมพันธ์ระหว่างมิติในระดับสูงขึ้น พบว่าวิธี NOP มีค่าการกระจายและค่ากลางน้อยกว่าวิธี M อย่างเห็นได้ชัด แสดงว่าวิธี NOP มีความคงที่มากกว่าวิธี M และข้อมูลส่วนใหญ่มีค่าความคลาดเคลื่อนของการประมาณค่าน้อยกว่าวิธี M นอกจากนี้พบว่าเมื่อความสัมพันธ์ระหว่างมิติอยู่ในระดับสูงในโครงสร้างที่มีความซับซ้อนมากขึ้น วิธี M จะมีกระจายของค่า RMSE สูงกว่าเงื่อนไขอื่น แสดงว่ามีความคงที่น้อยกว่ารูปแบบอื่น

โดยสรุปสำหรับค่าอำนาจจำแนกทั้ง 2 มิติ พบว่าลักษณะการกระจายของดัชนี RMSE มีลักษณะที่ไม่แตกต่างกันมากนัก โดยเมื่อความสัมพันธ์ระหว่างมิติความสามารถสูงขึ้นมีแนวโน้มที่จะมีค่าดัชนี RMSE รวมถึงค่ากลางเพิ่มขึ้นโดยเฉพาะวิธี M และผลการพิจารณาโดยภาพรวมพบว่าวิธี NOP มีความคงที่ในการประมาณค่าที่ดีกว่าวิธี M เมื่อพิจารณาค่ากลาง พบว่า วิธี NOP มีค่าน้อยกว่าวิธี M ในทุกเงื่อนไขที่ศึกษา ซึ่งมีลักษณะที่คล้ายคลึงกับการแปลงคะแนนจากระดับที่ 1 ไปยังระดับที่ 2

3) ดัชนี RMSE สำหรับค่าความยากของข้อสอบ (d)



แผนภาพที่ 4.24 แผนภูมิ Box Plot ของข้อมูลชุดวัดซ้ำทั้งหมด 20 ครั้ง สำหรับดัชนี RMSE สำหรับค่าพารามิเตอร์ d รายเงื่อนไขที่ศึกษา (L2_3)

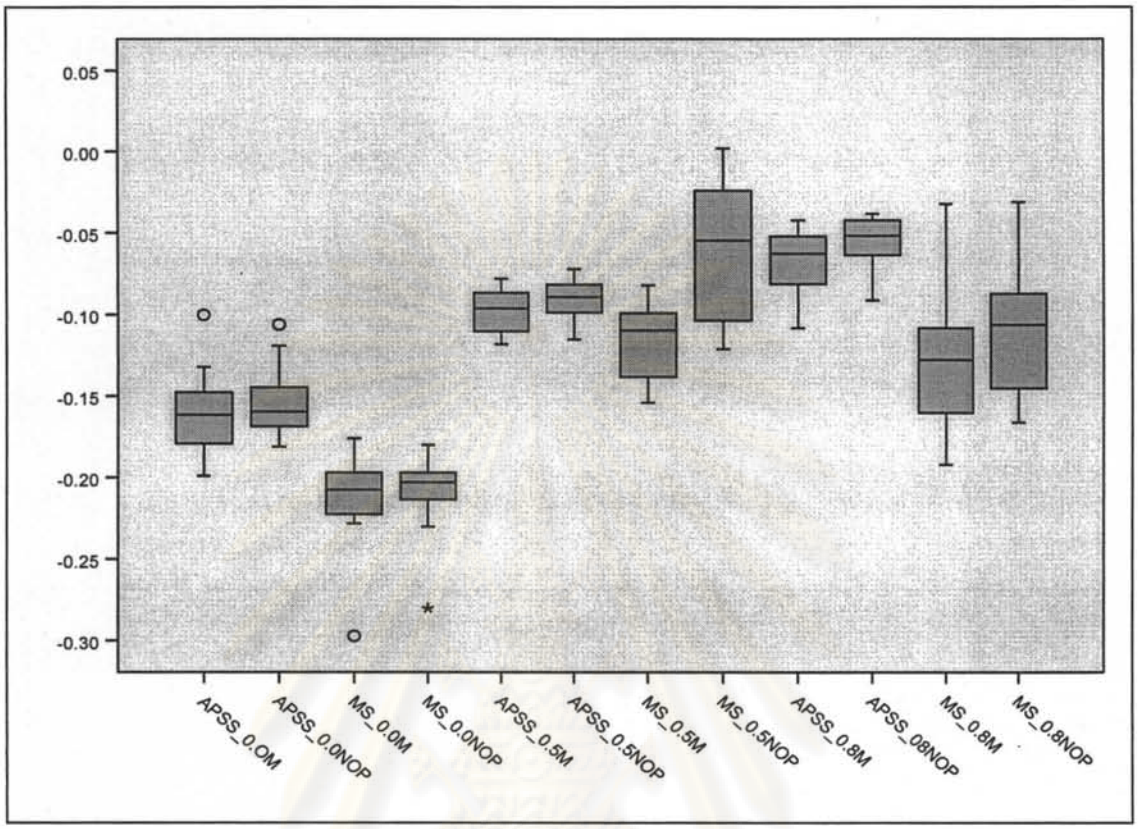
เมื่อพิจารณาดัชนี RMSE สำหรับค่าพารามิเตอร์ d พบว่า โครงสร้างมิติความสามารถที่ไม่ซับซ้อนมีช่วงของค่า RMSE สูงกว่าโครงสร้างที่ซับซ้อน โดยเมื่อความสัมพันธ์ระหว่างมิติเพิ่มสูงขึ้นจะมีค่า RMSE ลดลง นั่นคือมีความคลาดเคลื่อนของการประมาณค่าน้อยนั่นเอง โดยเมื่อความสัมพันธ์ในระดับปานกลางที่มีโครงสร้างซับซ้อนมากขึ้น พบว่ามีการกระจายของคะแนนน้อยกว่าเงื่อนไขอื่นอย่างเห็นได้ชัด แสดงว่ามีความคงที่ของดัชนี RMSE มากกว่า

2.2.2 ผลการวิเคราะห์โดยใช้ดัชนี BIAS เป็นตัวแปรตาม พบว่ามีเพียง 2 เงื่อนไขที่มีผลทำให้ค่าเฉลี่ยดัชนี BIAS ของวิธี M แตกต่างจากวิธี NOP อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 คือเงื่อนไข MS_0.5 ทั้งค่าพารามิเตอร์ a_1 (t=-4.840) และ a_2 (t=-5.567) และเงื่อนไข APSS_0.8 ทั้งค่าพารามิเตอร์ a_1 (t=-2.292) และ a_2 (t=-2.484) เช่นเดียวกัน โดยทั้งสองวิธีมีการประมาณค่าที่ต่ำกว่าความเป็นจริง แต่วิธี M จะมีการประมาณค่าที่ต่ำกว่าความเป็นจริงมากกว่าวิธี NOP อย่างเห็นได้ชัด

เมื่อพิจารณาโดยภาพรวม พบว่า โครงสร้างที่ซับซ้อนมากขึ้น (MS) มีการประมาณค่าต่ำกว่าความเป็นจริงสูงกว่าโครงสร้างที่ไม่ซับซ้อน (APSS) ในทุกระดับความสัมพันธ์สำหรับวิธี M

จากที่กล่าวมาสามารถแสดงแผนภาพ Box Plot เพื่อพิจารณาลักษณะการกระจายของค่าดัชนี BIAS ของพารามิเตอร์การแปลงคะแนนในมิติที่ 1 และ 2 ดังต่อไปนี้

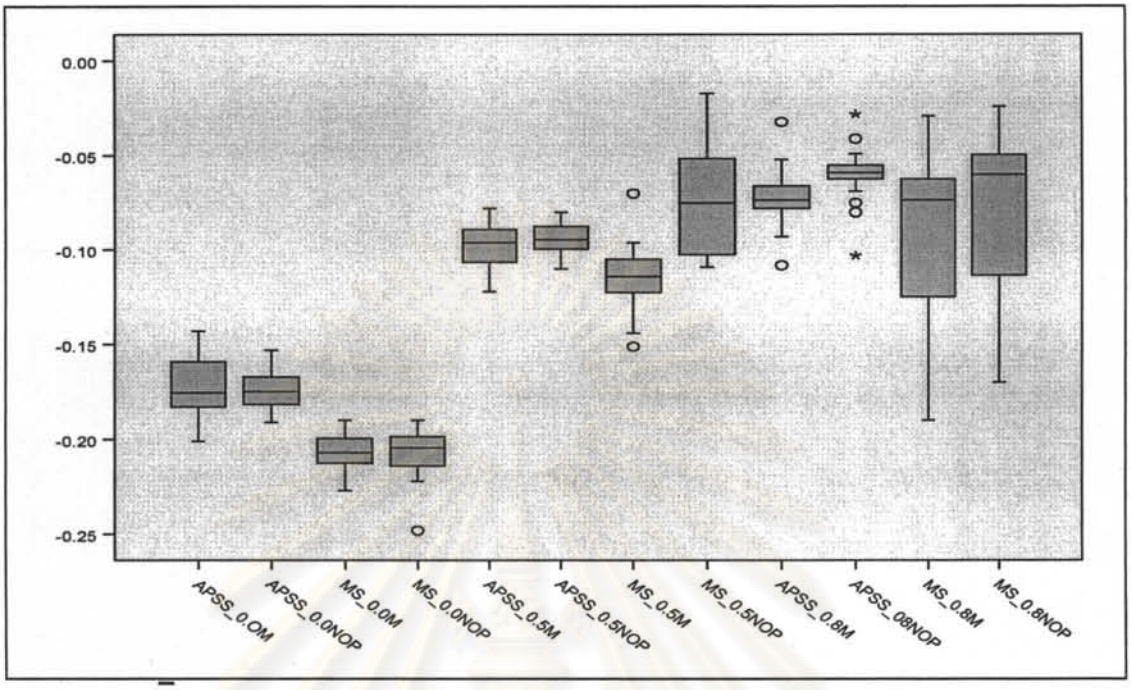
1) ดัชนี BIAS สำหรับค่าอำนาจจำแนกของข้อสอบในมิติที่ 1 (a_1)



แผนภาพที่ 4.25 แผนภูมิ Box Plot ของข้อมูลชุดวัดซ้ำทั้งหมด 20 ครั้ง สำหรับดัชนี BIAS ของค่าอำนาจจำแนกข้อสอบในมิติที่ 1 (a_1) รายเงื่อนไขที่ศึกษา (L2_3)

จากแผนภูมิ Box Plot แสดงให้เห็นถึงลักษณะของการกระจายค่า BIAS สำหรับอำนาจจำแนกในมิติที่ 1 จากการทำซ้ำ 20 ครั้ง เมื่อพิจารณารายเงื่อนไข พบว่าเมื่อความสัมพันธ์ระหว่างมิติอยู่ในระดับที่ไม่สูง การกระจายของวิธี M และ วิธี NOP ให้ค่าที่คล้ายคลึงกัน แต่เมื่อความสัมพันธ์ระหว่างมิติสูงขึ้นการกระจายของข้อมูลวิธีทั้งสองมีลักษณะแตกต่างกัน โดยวิธี M มีการกระจายของข้อมูลมากกว่าวิธี NOP เมื่อเปรียบเทียบระหว่างโครงสร้างมิติความสามารถ พบว่า โครงสร้างที่ซับซ้อนมากขึ้นมีแนวโน้มที่จะประมาณค่าต่ำกว่าความเป็นจริงมากกว่าโครงสร้างที่ไม่ซับซ้อน เมื่อพิจารณาโดยภาพรวมพบว่า การกระจายของทั้งสองวิธีจะประมาณค่าต่ำกว่าความเป็นจริงทั้งสองวิธี แต่วิธี M จะมีความคลาดเคลื่อนสูงกว่าวิธี NOP

2) ดัชนี BIAS สำหรับค่าอำนาจจำแนกของข้อสอบในมิติที่ 2 (a_2)

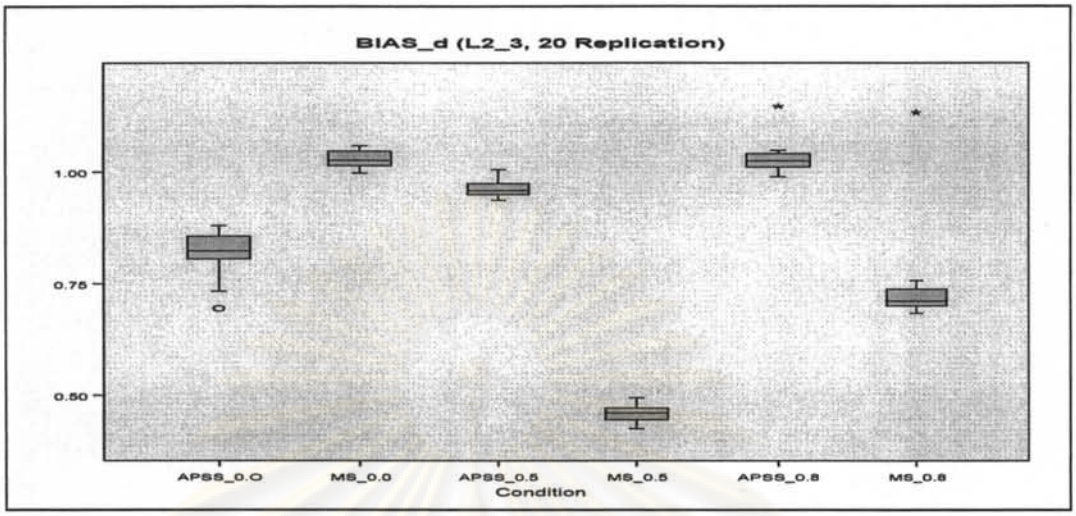


แผนภาพที่ 4.26 แผนภูมิ Box Plot ของข้อมูลชุดวัดซ้ำทั้งหมด 20 ครั้ง สำหรับดัชนี BIAS ของค่าอำนาจจำแนกของข้อสอบในมิติที่ 2 (a_2) รายเงื่อนไขที่ศึกษา (L2_3)

จากแผนภูมิ Box Plot เมื่อพิจารณารายเงื่อนไข พบว่าเมื่อความสัมพันธ์ระหว่างมิติเพิ่มขึ้นจะมีลักษณะของการกระจายค่า BIAS แตกต่างกันระหว่างสองวิธี ซึ่งโดยส่วนใหญ่วิธี NOP มีการกระจายของ BIAS น้อยกว่าวิธี M มีเพียงเงื่อนไข MS_0.5 ที่วิธี M มีการกระจายน้อยกว่าวิธี NOP แสดงว่าในเงื่อนไขดังกล่าววิธี M มีความคงที่ของดัชนี BIAS มากกว่าวิธี NOP เมื่อพิจารณาโดยภาพรวมพบว่า การกระจายของทั้งสองวิธีจะประมาณค่าต่ำกว่าความเป็นจริง แต่วิธี M จะมีความคลาดเคลื่อนสูงกว่าวิธี NOP อย่างชัดเจน อย่างไรก็ตามเมื่อความสัมพันธ์ระหว่างมิติอยู่ในระดับสูงสำหรับวิธี NOP จะมีการประมาณค่าใกล้เคียงกับความเป็นจริงมากกว่าเงื่อนไขอื่น โดยเมื่อความสัมพันธ์ระหว่างมิติความสามารถสูงขึ้นมีแนวโน้มในการประมาณค่าใกล้เคียงกับค่าที่แท้จริงมากกว่าเงื่อนไขที่ไม่มีความสัมพันธ์ระหว่างมิติอย่างชัดเจน

โดยสรุปสำหรับค่าอำนาจจำแนกทั้ง 2 มิติ พบว่าลักษณะการกระจายของดัชนี BIAS มีลักษณะที่ไม่แตกต่างกันมากนัก โดยเมื่อความสัมพันธ์ระหว่างมิติความสามารถสูงขึ้นมีแนวโน้มที่จะมีค่าดัชนี BIAS ลดลง ซึ่งโดยส่วนใหญ่วิธี M มีการประมาณค่าต่ำกว่าความเป็นจริงมากกว่าวิธี NOP ในขณะที่เมื่อความสัมพันธ์ระหว่างมิติสูงขึ้นวิธี NOP มีแนวโน้มในการประมาณค่ามากกว่าความเป็นจริง โดยภาพรวมพบว่ามีลักษณะที่ใกล้เคียงกับการแปลงจากระดับที่ 1 ไปยังระดับที่ 2

3) ดัชนี BIAS สำหรับค่าความยากของข้อสอบ (d)



แผนภาพที่ 4.27 แผนภูมิ Box Plot ของข้อมูลชุดวัดซ้ำทั้งหมด 20 ครั้ง สำหรับดัชนี BIAS สำหรับค่าพารามิเตอร์ d รายเงื่อนไขที่ศึกษา (L2_3)

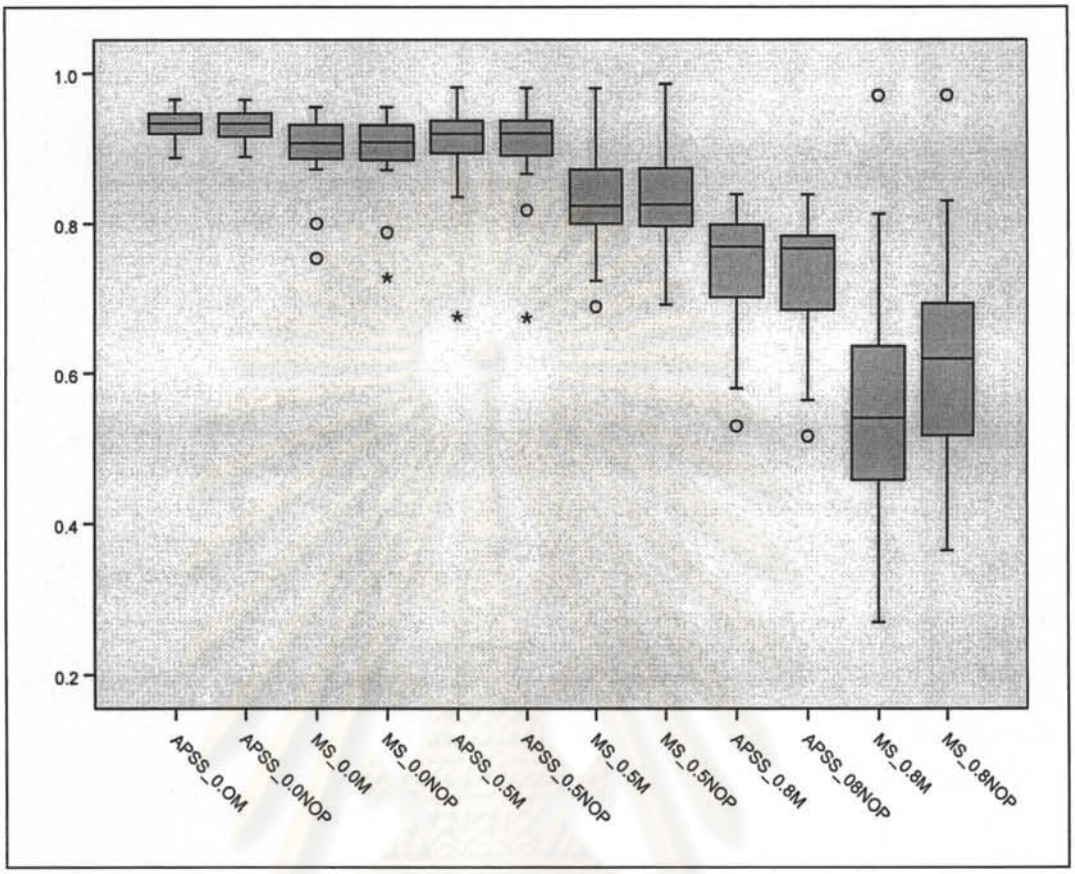
สำหรับค่าดัชนี BIAS พบว่า เมื่อไม่มีความสัมพันธ์ระหว่างมิติ โครงสร้างที่ไม่ซับซ้อนมีค่าดัชนี BIAS น้อยกว่า โครงสร้างที่ซับซ้อน แสดงว่ามีความถูกต้องในการแปลงคะแนนมากกว่า ในขณะที่เมื่อความสัมพันธ์ระหว่างมิติมีระดับสูงขึ้นพบว่า โครงสร้างที่ซับซ้อนมีค่า BIAS น้อยกว่าโครงสร้างที่ไม่ซับซ้อน ซึ่งโดยภาพรวมจะให้ค่าที่สูงกว่าค่าการเดาอย่างเห็นได้ชัดเนื่องจากมีค่ามากกว่าศูนย์

2.2.3) ผลการวิเคราะห์โดยใช้ค่า CORR เป็นตัวแปรตาม พบว่าไม่มีเงื่อนไขใดที่มีผลทำให้ค่าสัมประสิทธิ์สหสัมพันธ์ของวิธี M แตกต่างจากวิธี NOP อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 อย่างไรก็ตามแม้ผลการวิเคราะห์พบว่าทั้ง 2 วิธีให้ผลไม่แตกต่างกันทุกเงื่อนไข แต่ข้อค้นพบในการศึกษาจากค่า CORR ของ a_1 และ a_2 พบว่า โครงสร้างที่ไม่ซับซ้อนมีค่า CORR สูงกว่าในทุกระดับความสัมพันธ์ โดยเมื่อความสัมพันธ์ระหว่างมิติสูงขึ้นจะมีค่า CORR ลดน้อยลง

จากที่กล่าวมาสามารถแสดงแผนภาพ Box Plot เพื่อพิจารณาลักษณะการกระจายของค่า CORR ของพารามิเตอร์การแปลงคะแนนในมิติที่ 1 และ 2 ดังต่อไปนี้

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

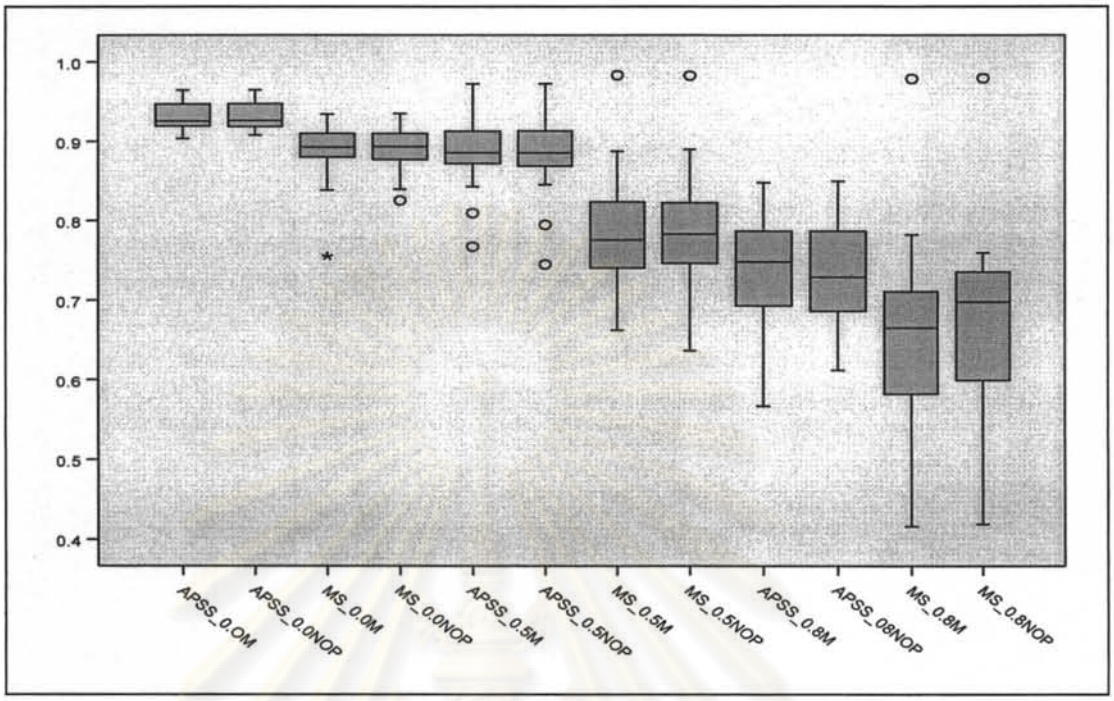
1) ค่า CORR สำหรับค่าอำนาจจำแนกของข้อสอบในมิติที่ 1 (a_1)



แผนภาพที่ 4.28 แผนภูมิ Box Plot ของข้อมูลชุดวัดซ้ำทั้งหมด 20 ครั้ง สำหรับค่า CORR ของค่าอำนาจจำแนกของข้อสอบในมิติที่ 1 (a_1) รายเงื่อนไขที่ศึกษา (L2_3)

จากแผนภาพที่ 4.28 แสดงให้เห็นถึงลักษณะของการกระจายค่า CORR สำหรับอำนาจจำแนกในมิติที่ 1 จากการทำซ้ำ 20 ครั้ง คล้ายคลึงกันทั้งสองวิธีในแต่ละเงื่อนไข โดยวิธี NOP มีช่วงการกระจายน้อยกว่าวิธี M เมื่อพิจารณาโดยภาพรวมพบว่าเมื่อความสัมพันธ์ระหว่างมิติเพิ่มสูงขึ้นมีแนวโน้มที่ค่า CORR ลดลง อย่างเห็นได้ชัดรวมถึงมีการกระจายของค่า CORR เพิ่มขึ้น ที่บ่งชี้ถึงความคงที่ของค่า CORR ลดลง เมื่อพิจารณาค่ากลางพบว่าทั้งสองวิธีมีค่า CORR ที่ใกล้เคียงกันแต่เมื่อระดับความสัมพันธ์ระหว่างมิติในระดับสูงสำหรับโครงสร้างที่ซับซ้อนมากขึ้น มีค่า CORR อยู่ในระดับที่ไม่สูงนัก โดยวิธี NOP มีค่าสูงกว่าวิธี M อย่างเห็นได้ชัด แสดงว่าความสัมพันธ์ระหว่างคะแนนแปลงกับคะแนนเกณฑ์โดยวิธี NOP มีค่าสูงกว่าวิธี M ซึ่งจากที่กล่าวมามีลักษณะที่ใกล้เคียงจากระดับที่ 1 ไปยังระดับที่ 2

2) ค่า CORR สำหรับค่าอำนาจจำแนกของข้อสอบในมิติที่ 2 (a_2)



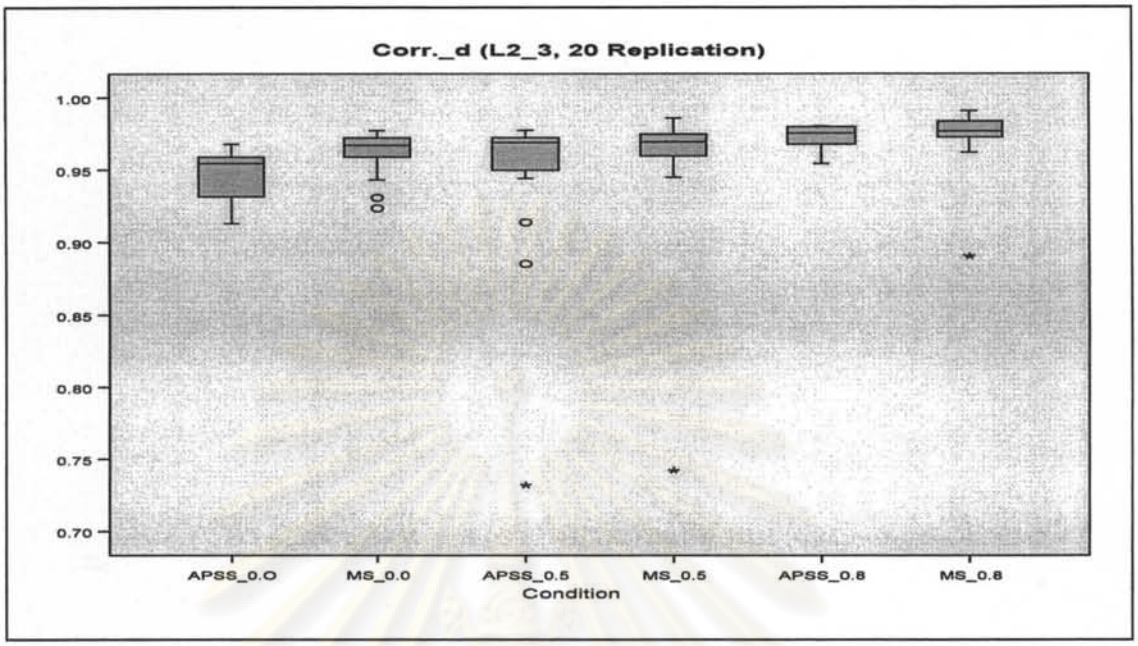
แผนภาพที่ 4.29 แผนภูมิ Box Plot ของข้อมูลชุดวัดซ้ำทั้งหมด 20 ครั้ง สำหรับค่า CORR ของค่าอำนาจจำแนกของข้อสอบในมิติที่ 2 (a_2) รายเงื่อนไขที่ศึกษา (L2_3)

จากแผนภาพที่ 4.29 เมื่อพิจารณารายเงื่อนไข พบว่าเมื่อความสัมพันธ์ระหว่างมิติอยู่ในระดับที่ไม่สูง จะมีการกระจายของข้อมูลน้อย ในขณะที่เมื่อความสัมพันธ์ระหว่างมิติอยู่ในระดับสูงจะมีช่วงการกระจายของข้อมูลมากขึ้นรวมถึงมีค่า CORR ลดลง อย่างไรก็ตาม โดยภาพรวมพบว่าวิธี NOP มีค่า COOR สูงกว่าวิธี NOP แสดงว่าความสัมพันธ์ระหว่างคะแนนแปลงกับคะแนนเกณฑ์โดยวิธี NOP มีค่าสูงกว่าวิธี M

โดยสรุปสำหรับค่าอำนาจจำแนกทั้ง 2 มิติ พบว่ามีลักษณะที่ไม่แตกต่างจากการแปลงคะแนนจากระดับที่ 1 ไปยังระดับที่ 2 นั่นคือ ลักษณะการกระจายของค่า CORR มีลักษณะที่ไม่แตกต่างกันมากนัก แสดงว่าความสัมพันธ์ระหว่างคะแนนแปลงกับคะแนนเกณฑ์ทั้งสองวิธีไม่แตกต่างกัน โดยเมื่อความสัมพันธ์ระหว่างมิติความสามารถสูงขึ้นมีแนวโน้มที่จะมีค่า CORR ลดลงอย่างเห็นได้ชัด

จุฬาลงกรณ์มหาวิทยาลัย

4.1.7 ค่า CORR สำหรับค่าความยากของข้อสอบ (d)



แผนภาพที่ 4.30 แผนภูมิ Box Plot ของข้อมูลชุดวัดซ้ำทั้งหมด 20 ครั้ง สำหรับค่า CORR ของพารามิเตอร์ d รายเงื่อนไขที่ศึกษา (L2_3)

เมื่อพิจารณาจากแผนภาพที่ 4.30 พบว่า เมื่อความสัมพันธ์ระหว่างมิติความสามารถเพิ่มขึ้นมีผลทำให้ค่า CORR เพิ่มสูงขึ้น โดยทุกเงื่อนไขมีค่าเข้าใกล้ 1

เมื่อพิจารณาโดยภาพรวมจากแผนภูมิ Box Plot ของค่าดัชนีบ่งชี้คุณภาพที่ได้จากการแปลงคะแนนจากระดับชั้นที่ 1 ไปยังระดับที่ 2 และจากระดับที่ 2 ไปยังระดับที่ 3 พบว่ามีลักษณะที่ใกล้เคียงกันเมื่อพิจารณาจากค่าอำนาจจำแนกของข้อสอบทั้งในมิติที่ 1 และมิติที่ 2 นั่นคือ เมื่อไม่มีความสัมพันธ์ระหว่างมิติพบว่าวิธี M และวิธี NOP มีการกระจายของดัชนี RMSE ใกล้เคียงกันทั้งวิธี M และ วิธี NOP แต่เมื่อความสัมพันธ์ระหว่างมิติเพิ่มมากขึ้นวิธี NOP มีความคลาดเคลื่อนของการประมาณค่าน้อยกว่าวิธี M เมื่อพิจารณาค่า BIAS พบว่าทั้งสองวิธีมีการประมาณค่าที่ต่ำกว่าค่าที่แท้จริง แต่เมื่อความสัมพันธ์ระหว่างมิติความสารถและโครงสร้างมีความซับซ้อนมากขึ้น มีแนวโน้มที่จะประมาณค่าสูงกว่าค่าที่แท้จริงมากกว่า อย่างไรก็ตามโดยภาพรวมพบว่าวิธี M มีการประมาณค่าที่สูงกว่าความเป็นจริงมากกว่าวิธี NOP สำหรับค่า CORR พบว่าเมื่อความสัมพันธ์ระหว่างมิติเพิ่มขึ้นมีผลทำให้ค่า CORR ลดลง โดยเฉพาะอย่างยิ่งในโครงสร้างที่มีความซับซ้อนเพิ่มขึ้น

สำหรับการพิจารณาค่าพารามิเตอร์ d พบว่าการแปลงคะแนนจากระดับชั้นที่ 1 ไปยังระดับที่ 2 มีความแตกต่างจากระดับที่ 2 ไปยังระดับที่ 3 นั่นคือการแปลงคะแนนจากระดับชั้นที่ 1 ไปยังระดับที่ 2 เมื่อความสัมพันธ์ระหว่างมิติเพิ่มสูงขึ้นจะมีค่า RMSE เพิ่มขึ้น แต่

เมื่อแปลงคะแนนจากระดับชั้นที่ 2 ไปยังระดับที่ 3 พบว่า เมื่อความสัมพันธ์ระหว่างมิติเพิ่มสูงขึ้นจะมีค่า RMSE ลดลง นั่นคือมีความคลาดเคลื่อนของการประมาณค่าน้อยนั่นเองสำหรับค่าดัชนี BIAS พบว่า เมื่อแปลงคะแนนจากระดับ 2 ไปยังระดับ 3 มีลักษณะที่แตกต่างจากการแปลงคะแนนจากระดับที่ 1 ไปยังระดับที่ 2 โดยเมื่อไม่มีความสัมพันธ์ระหว่างมิติพบว่าโครงสร้าง มีค่าดัชนี BIAS น้อยกว่า โครงสร้างที่ซับซ้อน ในขณะที่เมื่อความสัมพันธ์ระหว่างมิติมีระดับสูงขึ้นพบว่า โครงสร้างที่ซับซ้อนมีค่า BIAS น้อยกว่าโครงสร้างที่ไม่ซับซ้อน ซึ่งโดยภาพรวมจะให้ค่าที่สูงกว่าค่าการเดาอย่างเห็นได้ชัด เมื่อพิจารณาค่า CORR พบว่าเมื่อความสัมพันธ์ระหว่างมิติความสามารถเพิ่มขึ้นมีผลทำให้ค่า CORR เพิ่มสูงขึ้น โดยทุกเงื่อนไขมีค่าสัมประสิทธิ์สหสัมพันธ์เข้าใกล้ 1 ซึ่งมีลักษณะใกล้เคียงกับการแปลงคะแนนจากระดับที่ 1 ไปยังระดับที่ 2

โดยสรุป เมื่อศึกษาเชิงเปรียบเทียบระหว่างดัชนีบ่งชี้คุณภาพของวิธีการเชื่อมโยงคะแนนระหว่าง ขั้นที่ 1 จากระดับที่ 1 ไปยังระดับที่ 2 (L1_2) และ ขั้นที่ 2 จากระดับที่ 2 ไปยังระดับที่ 3 (L2_3) พบว่า ทั้ง 2 วิธี มีคุณภาพของการเชื่อมโยงคะแนนไม่แตกต่างกันมากนัก โดยเฉพาะความสัมพันธ์ระหว่างมิติมีค่าต่ำหรือไม่มีความสัมพันธ์ระหว่างมิติ แต่เมื่อความสัมพันธ์ระหว่างมิติเพิ่มขึ้น วิธี M มีความคลาดเคลื่อนของการประมาณค่าสูงกว่าวิธี NOP และมีความคลาดเคลื่อนของการประมาณค่าต่ำกว่าความเป็นจริงมากกว่าวิธี NOP ในขณะที่ความสัมพันธ์เกือบทุกเงื่อนไขมีค่าเข้าใกล้ 1 แต่มีแนวโน้มที่แตกต่างกันระหว่างค่าอำนาจจำแนกในมิติที่ 1 และมิติที่ 2 รวมถึงค่าความยากของข้อสอบ นั่นคือ สำหรับค่าอำนาจจำแนกเมื่อความสัมพันธ์ระหว่างมิติระดับสูงขึ้นค่าสัมประสิทธิ์สหสัมพันธ์จะมีค่าลดลงอย่างเห็นได้ชัด ในขณะที่ค่าพารามิเตอร์ d เมื่อความสัมพันธ์ระหว่างมิติระดับสูงขึ้นค่าสัมประสิทธิ์สหสัมพันธ์จะมีค่าสูงขึ้น

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตอนที่ 3 สรุปผลการวิเคราะห์ตามวัตถุประสงค์การวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์ 2 ข้อ คือ 1) เพื่อศึกษาปฏิสัมพันธ์ของวิธีการเชื่อมโยงคะแนน และเงื่อนไขที่ทำการศึกษา คือ โครงสร้างมิติความสามารถ และระดับความสัมพันธ์ระหว่างมิติความสามารถที่มีต่อคุณภาพของการเชื่อมโยงคะแนน และ 2) เพื่อเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนนตามโมเดล MIRT ที่มีลักษณะการหมุนแกนแตกต่างกัน 2 วิธี คือ วิธีการหมุนแกนแบบตั้งฉากโดยวิธี M (Min's Method) และวิธีการหมุนแกนแบบไม่ตั้งฉากโดยวิธี NOP (Non-Orthogonal Procrustes Method) ภายใต้โครงสร้างมิติความสามารถ (APSS และ MS) และระดับความสัมพันธ์ระหว่างมิติความสามารถ (r=0.0, r=0.5 และ r=0.8) ที่ต่างกัน

สำหรับคุณภาพของการเชื่อมโยงคะแนน พิจารณาจากการเปรียบเทียบ 3 ดัชนีบ่งชี้คุณภาพ คือ ***1) RMSE ซึ่งสะท้อนถึงความคงที่ (Stability)*** ของการเชื่อมโยงคะแนน โดยมีหลักเกณฑ์ในการพิจารณาจากการเปรียบเทียบค่าที่ได้ ถ้ามีค่าดัชนี RMSE น้อยกว่า แสดงว่ามีความคงที่ของการเชื่อมโยงคะแนนสูงกว่า สะท้อนให้เห็นถึงความคลาดเคลื่อนของการเชื่อมโยงคะแนนมีน้อย แต่ถ้ามีค่าดัชนี RMSE มากกว่า แสดงว่ามีความคงที่ของการเชื่อมโยงคะแนนต่ำ สะท้อนให้เห็นถึงความคลาดเคลื่อนของการเชื่อมโยงคะแนนมีมาก ***2) BIAS แสดงถึงความถูกต้อง (Accuracy)*** ของการเชื่อมโยงคะแนน โดยมีหลักเกณฑ์ในการพิจารณาจากการเปรียบเทียบค่าที่ได้ ถ้ามีค่าดัชนี BIAS ใกล้ศูนย์มากกว่าสะท้อนให้เห็นถึงการประมาณค่าใกล้เคียงกับค่าที่แท้จริงมากกว่า โดยเครื่องหมายที่ได้จะเป็นตัวสะท้อนถึงความคลาดเคลื่อนของการประมาณค่าที่ห่างจากค่าที่แท้จริงในลักษณะใด นั่นคือ ถ้าดัชนี BIAS ติดลบ (-) แสดงว่ามีการประมาณค่าต่ำกว่าความเป็นจริง (Underestimate) ถ้าเป็นบวก (+)แสดงว่ามีการประมาณค่าสูงกว่าความเป็นจริง (Overestimate) และ **3) *CORR แสดงถึงความสัมพันธ์ระหว่างค่าที่แปลงได้จากแบบสอบปรับเทียบคะแนนกับคะแนนเกณฑ์ที่ได้จากแบบสอบฐาน*** โดยมีหลักเกณฑ์ในการพิจารณาจากการเปรียบเทียบค่าที่ได้ ถ้ามีค่า CORR สูง จะสะท้อนให้เห็นถึงคุณภาพของการเชื่อมโยงคะแนนที่ดี

ผลการวิเคราะห์ข้อมูลตามวัตถุประสงค์การวิจัยสรุปได้ดังนี้

3.1 ผลการศึกษาปฏิสัมพันธ์ของวิธีการเชื่อมโยงคะแนน โครงสร้างมิติความสามารถ และระดับความสัมพันธ์ระหว่างมิติความสามารถที่ส่งผลต่อคุณภาพของการเชื่อมโยงคะแนน สามารถสรุปผลการทดสอบอิทธิพลระหว่างตัวแปรอิสระ (Test of Between-subject Effects) ของการแปลงคะแนนจากระดับ 1 ไปยังระดับ 2 (L1_2) และ การแปลงคะแนนจากระดับ 2 ไปยังระดับ 3 (L2_3) ดังตารางที่ 4.18

ตารางที่ 4.18 สรุปผลการทดสอบอิทธิพลระหว่างตัวแปรอิสระ (Test of Between-subject Effects) ของการแปลงคะแนนจากระดับ 1 ไปยังระดับ 2 (L1_2) และ การแปลงคะแนนจากระดับ 2 ไปยังระดับ 3 (L2_3)

แหล่งความแปรปรวน	ตัวแปรตาม	L1_2		L2_3	
		F	เปรียบเทียบรายคู่	F	เปรียบเทียบรายคู่
วิธีการเชื่อมโยงคะแนน(METHOD)	1. Ln RMSE_a1	*√	M>NOP*	x	M>NOP
	2. Ln RMSE_a2	**√	M>NOP**	x	M>NOP
	3. BIAS_a1	**√	M>NOP**	**√	M>NOP**
	4. BIAS_a2	**√	M>NOP**	**√	M>NOP**
	5. CORR_a1	x	NOP>M	x	NOP>M
	6. CORR_a2	x	NOP>M	x	NOP>M
โครงสร้างมิติความสามารถ (DS)	1. Ln RMSE_a1	x	MS>APSS	*√	MS>APSS*
	2. Ln RMSE_a2	**√	MS>APSS**	x	MS>APSS
	3. BIAS_a1	**√	MS>APSS**	**√	MS>APSS**
	4. BIAS_a2	x	MS>APSS	**√	MS>APSS**
	5. CORR_a1	**√	APSS>MS**	**√	M>NOP**
	6. CORR_a2	**√	APSS>MS**	**√	M>NOP**
ระดับความสัมพันธ์ระหว่างมิติความสามารถ (r)	1. Ln RMSE_a1	**√	0.8>0.0>0.5**	**√	0.8>0.0>0.5 0.8>0.0**
	2. Ln RMSE_a2	**√	0.8>0.0>0.5**	**√	0.8>0.0>0.5 0.8>0.0**
	3. BIAS_a1	**√	0.0>0.5>0.8**	**√	0.0>0.5>0.8 0.5>0.8**
	4. BIAS_a2	**√	0.0>0.5>0.8**	**√	0.0>0.5>0.8**
	5. CORR_a1	**√	0.0>0.5>0.8**	**√	0.0>0.5>0.8**
	6. CORR_a2	**√	0.0>0.5>0.8**	**√	0.0>0.5>0.8**
METHOD * DS	1. Ln RMSE_a1	x	-	x	-
	2. Ln RMSE_a2	x	-	x	-
	3. BIAS_a1	**√	-	*√	-
	4. BIAS_a2	**√	-	*√	-
	5. CORR_a1	x	-	x	-
	6. CORR_a2	x	-	x	-
METHOD * r	1. Ln RMSE_a1	x	-	x	-
	2. Ln RMSE_a2	*√	-	x	-
	3. BIAS_a1	*√	-	*√	-
	4. BIAS_a2	x	-	**√	-
	5. CORR_a1	x	-	x	-
	6. CORR_a2	x	-	x	-
DS * r	1. Ln RMSE_a1	**√	-	**√	-
	2. Ln RMSE_a2	**√	-	x	-
	3. BIAS_a1	**√	-	**√	-
	4. BIAS_a2	x	-	**√	-
	5. CORR_a1	**√	-	**√	-
	6. CORR_a2	**√	-	x	-
METHOD * DS * r	1. Ln RMSE_a1	x	-	x	-
	2. Ln RMSE_a2	x	-	x	-
	3. BIAS_a1	x	-	*√	-
	4. BIAS_a2	x	-	*√	-
	5. CORR_a1	*√	-	x	-
	6. CORR_a2	*√	-	x	-

* √ หมายถึง มีนัยสำคัญที่ .05 ** √ หมายถึง มีนัยสำคัญที่ .01 x หมายถึง ไม่มีนัยสำคัญทางสถิติ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

จากตารางที่ 4.18 สามารถสรุปในแต่ละรูปแบบของปฏิสัมพันธ์ และอิทธิพลหลักได้ดังนี้

3.1.1 ปฏิสัมพันธ์ระหว่าง 3 ตัวแปร (METHOD X DS X r)

เมื่อพิจารณาผลที่ได้จากการแปลงคะแนนจากระดับที่ 1 ไปยังระดับที่ 2 พบว่า วิธีการเชื่อมโยงคะแนน โครงสร้างมิติความสามารถ และระดับความสัมพันธ์ระหว่างมิติความสามารถ มีปฏิสัมพันธ์ที่ส่งผลต่อค่าความสัมพันธ์ระหว่างคะแนนที่แปลงได้กับคะแนนเกณฑ์ต่างกัน แต่ไม่ส่งผลต่อความคงที่และความถูกต้องของการเชื่อมโยงคะแนน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยมีปฏิสัมพันธ์ในทิศทางเดียวกัน (Ordinal Interation) นั่นคือ ในกรณีที่ชุดข้อสอบมีโครงสร้างที่ไม่ซับซ้อนผลของวิธีการเชื่อมโยงคะแนนและระดับความสัมพันธ์ระหว่างมิติความสามารถมีปฏิสัมพันธ์ร่วมกันในการส่งผลต่อค่า CORR โดยเมื่อความสัมพันธ์ระหว่างมิติความสามารถเท่ากับ 0 และ 0.5 พบว่าวิธี M และวิธี NOP มีค่าCORR ใกล้เคียงกัน แต่เมื่อระดับความสัมพันธ์หว่างมิติความสามารถเท่ากับ 0.8 พบว่า วิธี M จะมีค่า CORR มากกว่าวิธี NOP สำหรับกรณีที่แบบสอบมีโครงสร้างที่ซับซ้อนมากขึ้น (MS) พบว่าผลของวิธีการเชื่อมโยงคะแนนและระดับความสัมพันธ์ระหว่างมิติความสามารถมีปฏิสัมพันธ์ร่วมกันในการส่งผลต่อค่า CORR ของค่าโดยเมื่อความสัมพันธ์ระหว่างมิติความสามารถเท่ากับ 0 และ 0.5 พบว่าวิธี M และวิธี NOP มีค่า CORR ใกล้เคียงกัน เหมือนโครงสร้างที่ไม่ซับซ้อน (APSS) แต่ระดับความสัมพันธ์หว่างมิติความสามารถเท่ากับ 0.8 พบว่า วิธี NOP จะมีค่า CORR มากกว่าวิธี M

สำหรับผลการแปลงคะแนนจากระดับที่ 2 ไปยังระดับที่ 3 พบว่า วิธีการเชื่อมโยงคะแนน โครงสร้างมิติความสามารถ และระดับความสัมพันธ์ระหว่างมิติความสามารถมีปฏิสัมพันธ์ที่ส่งผลต่อความถูกต้องของการเชื่อมโยงคะแนนที่ต่างกัน แต่ไม่ส่งผลต่อความคงที่ค่าความสัมพันธ์ระหว่างคะแนนที่แปลงได้กับคะแนนเกณฑ์ของการเชื่อมโยงคะแนนอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยมีปฏิสัมพันธ์ในทิศทางเดียวกัน (Ordinal Interation) นั่นคือ

ในกรณีที่แบบสอบมีโครงสร้างที่ไม่ซับซ้อน (APSS) ผลของวิธีการเชื่อมโยงคะแนนและระดับความสัมพันธ์ระหว่างมิติความสามารถมีปฏิสัมพันธ์ร่วมกันในการส่งผลต่อค่าดัชนี BIAS ของค่าอำนาจจำแนกทั้ง 2 มิติ โดยในทุกระดับความสัมพันธ์วิธี M จะมีค่าดัชนี BIAS มากกว่าวิธี NOP ในทุกระดับความสัมพันธ์ แสดงให้เห็นว่า วิธี NOP มีการประมาณค่าใกล้เคียงกับค่าที่แท้จริงมากกว่าวิธี M ซึ่งสะท้อนให้เห็นถึงความถูกต้องในการเชื่อมโยงคะแนสำหรับ กรณีที่แบบสอบมีโครงสร้างที่ซับซ้อนมากขึ้น (MS) พบว่า ผลของวิธีการเชื่อมโยงคะแนนและระดับความสัมพันธ์ระหว่างมิติความสามารถมีปฏิสัมพันธ์ร่วมกันในการส่งผลต่อค่า BIAS ของค่าอำนาจจำแนกทั้ง 2 มิติ โดยในทุกระดับความสัมพันธ์วิธี M จะมี

ค่าดัชนี BIAS มากกว่าวิธี NOP แสดงให้เห็นว่า วิธี NOP มีการประมาณค่าใกล้เคียงกับค่าที่แท้จริงมากกว่าวิธี M โดยเมื่อระดับความสัมพันธ์อยู่ในระดับปานกลาง (r=0.5) จะมีค่า BIAS ต่ำกว่าระดับอื่น แสดงว่าเมื่อความสัมพันธ์อยู่ในระดับปานกลางจะมีการประมาณค่าใกล้เคียงกับค่าที่แท้จริงมากกว่าความสัมพันธ์ระหว่างมิติความสามารถในระดับอื่น

3.1.2 ปฏิสัมพันธ์ระหว่าง 2 ตัวแปร

ผลการศึกษาปฏิสัมพันธ์รายคู่ พบว่า ปฏิสัมพันธ์ระหว่างวิธีการเชื่อมโยงคะแนนและโครงสร้างมิติความสามารถ (METHOD X DS) และปฏิสัมพันธ์ระหว่างวิธีการเชื่อมโยงคะแนนและระดับความสัมพันธ์ระหว่างมิติความสามารถ (METHOD X r) ส่งผลต่อความคงที่และความถูกต้องของการเชื่อมโยงคะแนนที่ต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าคุณภาพของวิธีการเชื่อมโยงคะแนนที่ต่างกันขึ้นอยู่กับโครงสร้างมิติความสามารถและระดับความสัมพันธ์ระหว่างมิติความสามารถ แต่ไม่ส่งผลให้มีค่าความสัมพันธ์ระหว่างคะแนนที่แปลงได้กับคะแนนเกณฑ์ต่างกัน โดยมีปฏิสัมพันธ์ในทิศทางเดียวกัน (Ordinal Interation)

3.1.3 อิทธิพลหลัก

ผลการศึกษาอิทธิพลหลักโดยทำการเปรียบเทียบรายคู่ ปรากฏผลดังนี้

1) METHOD เมื่อเปรียบเทียบค่าเฉลี่ยของดัชนีบ่งชี้คุณภาพระหว่างวิธีการเชื่อมโยงคะแนน 2 วิธี โดยภาพรวมพบว่า วิธี NOP มีความคงที่และมีความถูกต้องในการประมาณค่ามากกว่าวิธี M แตกต่างกันอย่างมีนัยสำคัญทางสถิติ แต่ความสัมพันธ์ระหว่างค่าที่ได้จากการแปลงคะแนนกับคะแนนเกณฑ์ทั้งสองวิธีแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

2) DS เมื่อเปรียบเทียบค่าเฉลี่ยของดัชนีบ่งชี้คุณภาพระหว่างโครงสร้างมิติความสามารถระหว่างโครงสร้างที่ไม่ซับซ้อน (APSS) และโครงสร้างที่ซับซ้อนมากขึ้น (MS) โดยภาพรวม พบว่า โครงสร้างที่ไม่ซับซ้อน (APSS) มีความคงที่และมีความถูกต้องในการประมาณค่ามากกว่าโครงสร้างที่ซับซ้อนมากขึ้น (MS) โดยโครงสร้างที่ไม่ซับซ้อน (APSS) มีความสัมพันธ์ระหว่างค่าที่ได้จากจากการแปลงคะแนนกับคะแนนเกณฑ์ (CORR) สูงกว่าโครงสร้างที่ซับซ้อนมากขึ้น (MS)

3) เมื่อเปรียบเทียบค่าเฉลี่ยของดัชนีบ่งชี้คุณภาพระหว่างระดับความสัมพันธ์ระหว่างมิติความสามารถ (r) ในภาพรวมพบว่า เมื่อความสัมพันธ์เท่ากับ 0.5 จะมีความคงที่ในการประมาณค่าที่ดีกว่าระดับความสัมพันธ์ 0.0 และ 0.8 ตามลำดับ เมื่อพิจารณาถึงความถูกต้องในการประมาณค่า พบว่า เมื่อระดับความสัมพันธ์เท่ากับ 0.8 จะมีการประมาณค่าใกล้เคียงกับค่าที่แท้จริงมากกว่าระดับความสัมพันธ์ 0.5 และ 0.0 ตามลำดับ และเมื่อพิจารณาถึงค่า CORR ที่สะท้อนถึงความสัมพันธ์ระหว่างค่าที่ได้จากจากการแปลงคะแนนกับคะแนนเกณฑ์ พบว่าเมื่อความสัมพันธ์มีระดับสูงขึ้นจะมีค่า CORR ลดลง

3.2 สำหรับผลการเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนนตามโมเดล MIRT ระหว่างวิธี M และวิธี NOP ภายใต้โครงสร้างมิติความสามารถและระดับความสัมพันธ์ระหว่างมิติความสามารถที่ต่างกัน โดยพิจารณาจากการเปรียบเทียบดัชนี RMSE BIAS และ CORR โดยผลการวิจัยพบว่า ผลที่ได้จากการแปลงค่าพารามิเตอร์ความยากของข้อสอบ (d) ให้คุณภาพของการเชื่อมโยงคะแนนไม่แตกต่างกัน แต่ผลที่ได้จากการแปลงค่าพารามิเตอร์อำนาจจำแนกแตกต่างกันตามเงื่อนไขที่ศึกษา สามารถสรุปผลการเปรียบเทียบเป็น 2 ลักษณะ คือ

3.2.1 ผลการเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนนระหว่างวิธี M กับวิธี NOP ของแต่ละโครงสร้างมิติความสามารถ (APSS และ MS) และแต่ละระดับความสัมพันธ์มิติความสามารถ (r=0.0, r=0.5 และ r=0.8) ปรากฏผลดังตารางที่ 4.19

ตารางที่ 4.19 สรุปผลการทดสอบรายคู่ที่มีนัยสำคัญทางสถิติที่ระดับ .05

เงื่อนไขที่ศึกษา		RMSE				BIAS				CORR			
		a_1		a_2		a_1		a_2		a_1		a_2	
		L1_2	L2_3	L1_2	L2_3	L1_2	L2_3	L1_2	L2_3	L1_2	L2_3	L1_2	L2_3
โครงสร้างมิติ	APSS	x	x	x	x	x	x	x	x	x	x	x	x
ความสามารถ (DS)	MS	M>NOP	x	M>NOP	x	x	M>NOP	M>NOP	x	x	x	x	x
ระดับความสัมพันธ์	r=0.0	x	x	x	x	x	x	M>NOP	x	x	x	x	x
ระหว่างมิติ	r=0.5	x	x	x	x	M>NOP	M>NOP	M>NOP	M>NOP	x	x	x	x
ความสามารถ (r)	r=0.8	M>NOP	x	M>NOP	x	M>NOP	x	M>NOP	x	x	x	x	x

X ไม่พบว่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติ

จากตารางที่ 4.19 สามารถสรุปได้ การเชื่อมโยงคะแนนสำหรับชุดของข้อสอบที่มีโครงสร้างไม่ซับซ้อน (APSS) พบว่าทั้ง 2 วิธี มีคุณภาพของการเชื่อมโยงคะแนนไม่แตกต่างกัน แต่เมื่อโครงสร้างมีความซับซ้อนมากขึ้น พบว่าวิธี M มีค่าดัชนี RMSE และ BIAS มากกว่าวิธี NOP สะท้อนให้เห็นว่าวิธี NOP มีความคงที่และมีความถูกต้องของการเชื่อมโยงคะแนนมากกว่าวิธี M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

เมื่อพิจารณาระดับความสัมพันธ์ระดับมิติความสามารถ พบว่า ในทุกระดับความสัมพันธ์ระหว่างมิติความสามารถวิธี NOP จะมีค่า BIAS ต่ำกว่าวิธี M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 สะท้อนให้เห็นว่าวิธี NOP มีความถูกต้องในการประมาณค่ามากกว่าวิธี M อย่างไรก็ตามจากผลการวิจัยพบว่าระดับความสัมพันธ์ระหว่างมิติความสามารถจะส่งผลกระทบน้อยมากต่อค่า RMSE ซึ่งสะท้อนถึงความคงที่ของการเชื่อมโยงคะแนน อย่างไรก็ตามพบว่าเมื่อความสัมพันธ์

ระหว่างมิติความสามารถสูง (r=0.8) วิธี NOP จะมีค่า RMSE ต่ำกว่าวิธี M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าวิธี NOP มีความคงที่ในการเชื่อมโยงคะแนนมากกว่าวิธี M

ข้อสังเกตที่ได้จากผลการวิจัยครั้งนี้พบว่า ค่า CORR ทั้ง 2 วิธี ไม่ต่างกัน ไม่ว่าจะศึกษาจากโครงสร้างมิติความสามารถลักษณะใด และ ระดับความสัมพันธ์ระหว่างมิติความสามารถระดับใด แสดงว่าทั้งสองวิธีมีความสัมพันธ์ระหว่างค่าที่แปลงได้จากแบบสอบปรับเทียบคะแนนกับคะแนนเกณฑ์ที่ได้จากแบบสอบฐานไม่ต่างกัน

3.2.2 ผลการเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนนระหว่างวิธี M กับวิธี NOP ภายใต้การศึกษาข้ามเงื่อนไข ประกอบด้วย 6 เงื่อนไข คือ ภายใต้การศึกษาใน 6 เงื่อนไข คือ APSS_0.0, MS_0.0, APSS_0.5, MS_0.5, APSS_0.8 และ MS_0.8 สามารถสรุปวิธีการเชื่อมโยงคะแนนที่มีค่าดัชนี RMSE ต่ำ ดัชนี BIAS ต่ำ และค่า CORR สูง ในภาพรวมทุกสถานการณ์ได้ดังตารางที่ 4.20

ตารางที่ 4.20 สรุปวิธีการเชื่อมโยงคะแนนที่มีค่าดัชนี RMSE น้อย BIAS ต่ำ และมีค่า CORR สูง ในภาพรวมทุกเงื่อนไข เมื่อเปรียบเทียบระหว่างวิธี M กับวิธี NOP

เงื่อนไขการจำลองข้อมูล	RMSE ต่ำ				BIAS ต่ำ				CORR สูง			
	a1		a2		a1		a2		a1		a2	
	L1_2	L2_3	L1_2	L2_3	L1_2	L2_3	L1_2	L2_3	L1_2	L2_3	L1_2	L2_3
APSS_0.0	NOP	NOP	M ≈ NOP	M ≈ NOP	M	M	NOP	NOP	NOP	M =NOP	M ≈ NOP	M ≈ NOP
MS_0.0	NOP	NOP	NOP	NOP	**NOP****	NOP	**NOP****	M =NOP	NOP	M	M	NOP
APSS_0.5	NOP	NOP	NOP	NOP	**NOP****	NOP	**NOP****	NOP	NOP	M	M	M
MS_0.5	NOP	NOP	NOP	NOP	**NOP****	****NOP**	**NOP****	**NOP****	M	NOP	M	NOP
APSS_0.8	NOP	**NOP****	NOP	NOP	NOP	***NOP**	NOP	**NOP***	M	M	M	M
MS_0.8	**NOP****	NOP	**NOP****	NOP	**NOP****	NOP	**NOP****	NOP	**NOP***	NOP	**NOP***	NOP

หมายเหตุ *p<.05, **p<.01

M ≈NOP หมายถึง วิธี M มีค่าใกล้เคียงวิธี NOP

M = NOP หมายถึง วิธี M มีค่าเท่ากับวิธี NOP

จากตารางที่ 4.20 เมื่อพิจารณาจากดัชนี RMSE ที่สะท้อนถึงความคงที่ของการเชื่อมโยงคะแนนพบว่าวิธี NOP มีความคงที่มากกว่าวิธี M เกือบทุกสถานการณ์ที่ศึกษาโดยเฉพาะอย่างยิ่ง เงื่อนไขความสัมพันธ์ระหว่างมิติความสามารถอยู่ในระดับสูง (APSS_0.8, MS_0.8) เนื่องจากทั้งสองวิธีมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 อย่างไรก็ตามจากผลการศึกษาเงื่อนไขที่ไม่มีความสัมพันธ์ระหว่างมิติและความสัมพันธ์ระหว่างมิติไม่สูง พบว่าความคลาดเคลื่อนของการประมาณค่าการแปลงคะแนนไม่แตกต่างกันมากนัก แต่เมื่อความสัมพันธ์ระหว่างมิติเพิ่มขึ้น วิธี NOP มีความคงที่ของการเชื่อมโยงคะแนนมากกว่าวิธี M

เมื่อพิจารณาจากดัชนี BIAS ที่สะท้อนถึงความถูกต้องของการเชื่อมโยงคะแนน พบว่า วิธี NOP มีความถูกต้องของการเชื่อมโยงคะแนนมากกว่าวิธี M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เกือบทุกเงื่อนไข แต่อย่างไรก็ตามทั้งสองวิธีต่างมีการประมาณค่าที่ต่ำกว่าความเป็นจริง (Underestimate) โดยวิธี M จะมีการประมาณค่าที่ต่ำกว่าความเป็นจริงมากกว่าวิธี NOP อย่างเห็นได้ชัด

เมื่อพิจารณาค่า CORR ที่สะท้อนถึงความสัมพันธ์ระหว่างคะแนนที่แปลงได้กับคะแนนเกณฑ์ พบว่าเมื่อไม่มีความสัมพันธ์ระหว่างมิติหรือความสัมพันธ์ระหว่างมิติในระดับปานกลางวิธี M มีค่า CORR สูงกว่าวิธี NOP อย่างไรก็ตามพบว่าเมื่อความสัมพันธ์ระหว่างมิติสูงขึ้นวิธี NOP จะมีค่า CORR สูงกว่าวิธี M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 (MS_0.8) แต่ให้ค่าสัมประสิทธิ์สหสัมพันธ์ในระดับปานกลาง

เมื่อพิจารณาคุณภาพของการเชื่อมโยงคะแนนโดยรวมของวิธีการเชื่อมโยงคะแนนจากระดับที่ 1 ไประดับที่ 2 (L1_2) และจากระดับที่ 2 ไประดับที่ 3 (L2_3) พบว่าวิธีการเชื่อมโยงคะแนนทั้ง 2 วิธีมีคุณภาพของการเชื่อมโยงคะแนนไม่แตกต่างกันมากนัก เมื่ออยู่ในเงื่อนไขความสัมพันธ์ระหว่างมิติมีค่าต่ำหรือไม่มีความสัมพันธ์ระหว่างมิติ แต่เมื่อความสัมพันธ์ระหว่างมิติเพิ่มขึ้น วิธี NOP มีความคงที่ มีความถูกต้อง และมีค่าความสัมพันธ์ระหว่างคะแนนที่แปลงได้กับคะแนนเกณฑ์ของการเชื่อมโยงคะแนนมากว่าวิธี M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ซึ่งวิธี M มีแนวโน้มในการประมาณค่าที่ต่ำกว่าความเป็นจริง (Underestimate) มากกว่าวิธี NOP

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 5
สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

การวิจัยครั้งนี้เป็นการวิจัยเชิงทดลองโดยใช้วิธีการจำลองข้อมูล มีวัตถุประสงค์ 2 ประการ คือ 1) เพื่อศึกษาปฏิสัมพันธ์ของวิธีการเชื่อมโยงคะแนน โครงสร้างมิติความสามารถ และระดับความสัมพันธ์ระหว่างมิติความสามารถ และ 2) เพื่อเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนนตามโมเดล MIRT ที่มีลักษณะการหมุนแกนแตกต่างกัน 2 วิธี คือ วิธีการหมุนแกนแบบตั้งฉากโดยวิธี M (Min's Method) และวิธีการหมุนแกนแบบไม่ตั้งฉากโดยวิธี NOP (Non-Orthogonal Procrustes Method) ภายใต้โครงสร้างมิติความสามารถและระดับความสัมพันธ์ระหว่างมิติความสามารถที่แตกต่างกัน เงื่อนไขที่ทำการศึกษามี 12 เงื่อนไข (2X2X3) ประกอบด้วย วิธีการเชื่อมโยงคะแนน 2 วิธี คือ วิธี M และ วิธี NOP โครงสร้างมิติความสามารถ 2 ลักษณะ คือ โครงสร้างที่ไม่ซับซ้อน และโครงสร้างที่ซับซ้อนมากขึ้น และระดับความสัมพันธ์ระหว่างมิติความสามารถ 3 ระดับ คือ ไม่มีความสัมพันธ์ (r=0.0) มีความสัมพันธ์ระดับกลาง (r=0.5) และมีความสัมพันธ์ระดับสูง (r=0.8) ซึ่งมีการทำซ้ำเงื่อนไขละ 20 รอบ

โมเดลการวัดที่ใช้ในการวิจัยครั้งนี้เป็นโมเดลโลจิสติกแบบพหุมิติ ชนิด 3 พารามิเตอร์ (Multidimensional 3-Parameter Logistic Model; M3PL) ซึ่งเป็นโมเดล MIRT ที่การประมาณค่าความสามารถ (θ) ในมิติต่างกันสามารถชดเชยกันได้ (Compensatory Model) โดยมีการตอบสนองข้อสอบที่มีการตรวจให้คะแนน 2 ค่า และใช้เทคนิควิธีการสร้างสเกลแนวตั้งแบบพหุมิติ (Multidimensional Vertical Scaling) ตามโมเดล MIRT จากผู้สอบที่มีความสามารถแตกต่างกัน 3 ระดับชั้น ออกแบบการเก็บข้อมูลสำหรับผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้แบบสอบร่วม (Anchor-Test Nonequivalent Group Design) โดยนักเรียนในแต่ละระดับชั้นจะทำแบบสอบในระดับชั้นของตนเองและแบบสอบในระดับชั้นสูงกว่าที่อยู่ติดกัน ซึ่งกำหนดให้นักเรียนในระดับที่ 1 (n=2,000 คน) ทำแบบสอบระดับที่ 2 ส่วนนักเรียนในระดับที่ 2 (n=2,000 คน) ทำแบบสอบในระดับตนเองและทำแบบสอบระดับที่ 3 ส่วนนักเรียนระดับที่ 3 ทำแบบสอบในระดับชั้นของตนเอง ซึ่งมีข้อสอบร่วมชนิดภายใน (Internal-Common Items) จำนวน 30 ข้อ หรือ 30% ของจำนวนข้อสอบทั้งหมด 100 ข้อ

การจำลองข้อมูลมี 3 ขั้นตอน คือ 1) การสร้างค่าพารามิเตอร์ที่แท้จริงของข้อสอบและแบบแผนการตอบของนักเรียนแต่ละระดับ โดยใช้โปรแกรม MATLAB 2) การประมาณค่าพารามิเตอร์ตามโมเดล MIRT โดยใช้โปรแกรม TESTFACT และ BILOG-MG และ 3) การเชื่อมโยงคะแนนหรือการปรับเทียบคะแนน โดยใช้โปรแกรม MATLAB

เกณฑ์ที่ใช้ในการเปรียบเทียบคุณภาพวิธีการเชื่อมโยงคะแนนพิจารณาจากดัชนีค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) เพื่อสะท้อนถึงความคงที่ของการเชื่อมโยงคะแนน ค่าความลำเอียง (BIAS) เพื่อสะท้อนถึงความถูกต้องของการเชื่อมโยงคะแนน และค่าสัมประสิทธิ์สหสัมพันธ์ของค่าพารามิเตอร์จากแบบสอบฐานและค่าพารามิเตอร์ที่แปลงได้จากแบบสอบปรับเทียบคะแนน (CORR) เพื่อศึกษาความสัมพันธ์ระหว่างคะแนนที่แปลงได้กับคะแนนเกณฑ์ พร้อมทั้งวิเคราะห์ความแปรปรวนแบบพหุจำแนก 3 ทาง (Three-way MANOVA) เพื่อเปรียบเทียบค่าเฉลี่ยของดัชนี RMSE BIAS และค่า CORR สำหรับศึกษาปฏิสัมพันธ์ของเงื่อนไขที่ศึกษา โดยใช้โปรแกรม SPSS for Window Version 14.0 เพื่อประโยชน์ในการนำไปเลือกใช้กับสถานการณ์ต่าง ๆ ให้เหมาะสมกับข้อมูลจริงต่อไปในอนาคต

ในบทนี้ผู้วิจัยนำเสนอสรุปและอภิปรายผลการวิจัย รวมทั้งข้อเสนอแนะ ดังต่อไปนี้

สรุปผลการวิจัย

5.1 ผลการศึกษาปฏิสัมพันธ์ของวิธีการเชื่อมโยงคะแนน (METHOD) โครงสร้างมิติความสามารถ (DS) และระดับความสัมพันธ์ระหว่างมิติความสามารถ (r) เมื่อพิจารณาผลการแปลงค่าพารามิเตอร์จากระดับที่ 1 ไปยังระดับที่ 2 พบว่า METHOD X DSX r มีปฏิสัมพันธ์ที่ส่งผลต่อค่าความสัมพันธ์ระหว่างคะแนนที่แปลงได้กับคะแนนเกณฑ์ (CORR) ต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่ไม่ส่งผลต่อความคงที่และความถูกต้องของการเชื่อมโยงคะแนน โดยปฏิสัมพันธ์มีลักษณะเป็นไปในทิศทางเดียวกัน (Ordinal Interaction) นั่นคือ ในกรณีที่ชุดข้อสอบมีโครงสร้างที่ไม่ซับซ้อน ผลของวิธีการเชื่อมโยงคะแนนและระดับความสัมพันธ์ระหว่างมิติความสามารถมีปฏิสัมพันธ์ร่วมกันในการส่งผลต่อค่า CORR โดยเมื่อความสัมพันธ์ระหว่างมิติความสามารถเท่ากับ 0 และ 0.5 พบว่าวิธี M และวิธี NOP มีค่าCORR ใกล้เคียงกัน แต่เมื่อระดับความสัมพันธ์หว่างมิติความสามารถเท่ากับ 0.8 พบว่า วิธี M จะมีค่า CORR มากกว่าวิธี NOP สำหรับกรณีที่แบบสอบมีโครงสร้างที่ซับซ้อนมากขึ้น (MS) พบว่าผลของวิธีการเชื่อมโยงคะแนนและระดับความสัมพันธ์ระหว่างมิติความสามารถมีปฏิสัมพันธ์ร่วมกันในการส่งผลต่อค่า CORR โดยเมื่อความสัมพันธ์ระหว่างมิติความสามารถเท่ากับ 0 และ 0.5 พบว่าวิธี M และวิธี NOP มีค่า CORR ใกล้เคียงกัน เหมือนโครงสร้างที่ไม่ซับซ้อน (APSS) แต่ระดับความสัมพันธ์หว่างมิติความสามารถเท่ากับ 0.8 พบว่า วิธี NOP จะมีค่า CORR มากกว่าวิธี M

สำหรับผลการแปลงคะแนนจากระดับที่ 2 ไปยังระดับที่ 3 พบว่า METHOD X DSX r มีปฏิสัมพันธ์ซึ่งส่งผลต่อความถูกต้องของการเชื่อมโยงคะแนน (BIAS) ที่ต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่ไม่ส่งผลต่อความคงที่ (RMSE) และค่าความสัมพันธ์ระหว่างคะแนนที่แปลงได้กับคะแนนเกณฑ์ (CORR) ของการเชื่อมโยงคะแนน โดยมีปฏิสัมพันธ์

ในทิศทางเดียวกัน (Ordinal Interaction) นั่นคือ ในกรณีที่ชุดข้อสอบมีโครงสร้างไม่ซับซ้อน (APSS) ผลของวิธีการเชื่อมโยงคะแนนและระดับความสัมพันธ์ระหว่างมิติความสามารถมีปฏิสัมพันธ์ร่วมกันในการส่งผลต่อความถูกต้องของการเชื่อมโยงคะแนน โดยในทุกระดับความสัมพันธ์ วิธี NOP จะมีความถูกต้องในการเชื่อมโยงคะแนน สำหรับ กรณีที่แบบสอบมีโครงสร้างที่ซับซ้อนมากขึ้น (MS) พบว่า ผลของวิธีการเชื่อมโยงคะแนนและระดับความสัมพันธ์ระหว่างมิติความสามารถมีปฏิสัมพันธ์ร่วมกันในการส่งผลต่อความถูกต้องของการเชื่อมโยงคะแนน โดยในทุกระดับความสัมพันธ์ วิธี NOP จะมีความถูกต้องในการเชื่อมโยงคะแนนเช่นเดียวกัน แต่เมื่อความสัมพันธ์อยู่ในระดับปานกลางจะมีการประมาณค่าใกล้เคียงกับค่าที่แท้จริงมากกว่าความสัมพันธ์ระหว่างมิติความสามารถในระดับอื่น

ส่วนปฏิสัมพันธ์รายคู่ พบว่า METHOD x DS และ METHOD x DS มีปฏิสัมพันธ์ ซึ่งส่งผลต่อความคงที่และความถูกต้องของการเชื่อมโยงคะแนนที่ต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าคุณภาพของวิธีการเชื่อมโยงคะแนนที่ต่างกันขึ้นอยู่กับ 1) โครงสร้างมิติความสามารถและ 2) ระดับความสัมพันธ์ระหว่างมิติความสามารถ โดยมีปฏิสัมพันธ์ในทิศทางเดียวกัน แต่ไม่ส่งผลให้มีค่าความสัมพันธ์ระหว่างคะแนนที่แปลงได้กับคะแนนเกณฑ์ (CORR) ต่างกัน

สำหรับการศึกษา*อิทธิพลหลัก METHOD* พบว่า วิธี NOP มีความคงที่และมีความถูกต้องในการประมาณค่ามากกว่าวิธี M แตกต่างกันอย่างมีนัยสำคัญทางสถิติ แต่ความสัมพันธ์ระหว่างค่าที่ได้จากการแปลงคะแนนกับคะแนนเกณฑ์ทั้งสองวิธีแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ *ส่วนอิทธิพลหลัก DS* พบว่า โครงสร้างที่ไม่ซับซ้อน (APSS) มีความคงที่และมีความถูกต้องในการประมาณค่ามากกว่าโครงสร้างที่ซับซ้อนมากขึ้น (MS) โดยโครงสร้างที่ไม่ซับซ้อน (APSS) มีความสัมพันธ์ระหว่างค่าที่ได้จากจากการแปลงคะแนนกับคะแนนเกณฑ์ (CORR) สูงกว่าโครงสร้างที่ซับซ้อนมากขึ้น (MS) *และอิทธิพลหลัก r* พบว่าเมื่อความสัมพันธ์เท่ากับ 0.5 จะมีความคงที่ในการประมาณค่าที่ดีกว่าระดับความสัมพันธ์ 0.0 และ 0.8 ตามลำดับ เมื่อพิจารณาถึงความถูกต้องในการประมาณค่า พบว่า เมื่อระดับความสัมพันธ์เท่ากับ 0.8 จะมีการประมาณค่าใกล้เคียงกับค่าที่แท้จริงมากกว่าระดับความสัมพันธ์ 0.5 และ 0.0 ตามลำดับ และเมื่อพิจารณาถึงค่า CORR ที่สะท้อนถึงความสัมพันธ์ระหว่างค่าที่ได้จากการแปลงคะแนนกับคะแนนเกณฑ์ พบว่าเมื่อความสัมพันธ์มีระดับสูงขึ้นจะมีค่า CORR ลดลง

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

5.2 สำหรับผลการเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนนตามโมเดล MIRT ระหว่างวิธี M และวิธี NOP ภายใต้โครงสร้างมิติความสามารถและระดับความสัมพันธ์ระหว่างมิติความสามารถที่ต่างกัน พบว่า ผลที่ได้จากการแปลงค่าพารามิเตอร์ความยากของข้อสอบ (d) ให้คุณภาพของการเชื่อมโยงคะแนนไม่ต่างกัน แต่ผลที่ได้จากการแปลงค่าพารามิเตอร์อำนาจจำแนกแตกต่างกันตามเงื่อนไขที่ศึกษา สามารถสรุปได้ดังนี้

เมื่อพิจารณาเงื่อนไขภายใต้โครงสร้างมิติความสามารถ พบว่า การเชื่อมโยงคะแนนสำหรับชุดของข้อสอบที่มีโครงสร้างไม่ซับซ้อน (APSS) มีคุณภาพของการเชื่อมโยงคะแนนไม่แตกต่างกันทั้ง 2 วิธี แต่เมื่อโครงสร้างมีความซับซ้อนมากขึ้น พบว่า วิธี NOP มีความคงที่และมีความถูกต้องของการเชื่อมโยงคะแนนมากกว่าวิธี M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

เมื่อพิจารณาเงื่อนไขระดับความสัมพันธ์ระหว่างมิติความสามารถ พบว่า ในทุกระดับความสัมพันธ์วิธี NOP มีความถูกต้องในการประมาณค่ามากกว่าวิธี M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่ระดับความสัมพันธ์ระหว่างมิติความสามารถจะส่งผลกระทบน้อยต่อความคงที่ของการเชื่อมโยงคะแนน และจะส่งผลต่อความคงที่ของการเชื่อมโยงคะแนนเมื่อความสัมพันธ์ระหว่างมิติความสามารถสูง (r=0.8) โดย วิธี NOP มีความคงที่ในการเชื่อมโยงคะแนนมากกว่าวิธี M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 อย่างไรก็ตามผลการวิจัยพบว่าทั้งสองวิธีมีความสัมพันธ์ระหว่างค่าที่แปลงได้จากแบบสอบปรับเทียบคะแนนกับคะแนนเกณฑ์ที่ได้จากแบบสอบฐานไม่ต่างกันไม่ว่าจะศึกษาจากโครงสร้างมิติความสามารถลักษณะใดและระดับความสัมพันธ์ระหว่างมิติความสามารถระดับใด แสดงว่าทั้งสองวิธี

เมื่อพิจารณาคุณภาพของการเชื่อมโยงคะแนนโดยรวม พบว่า คุณภาพของการเชื่อมโยงคะแนนระหว่างวิธี M และวิธี NOP มีค่าใกล้เคียงกัน โดยเฉพาะเงื่อนไขที่ไม่มีความสัมพันธ์ระหว่างมิติความสามารถและมีความสัมพันธ์ระหว่างมิติความสามารถในระดับกลางกับโครงสร้างที่ไม่ซับซ้อน แต่เมื่อความสัมพันธ์ระหว่างมิติความสามารถเพิ่มขึ้นกับโครงสร้างที่ซับซ้อนมากขึ้น พบว่า วิธี NOP มีคุณภาพของการเชื่อมโยงคะแนนมากกว่าวิธี M อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 นั่นคือ วิธี NOP มีความคงที่ (RMSE) และความถูกต้อง (BIAS) ของการเชื่อมโยงคะแนนมากกว่าวิธี M ซึ่งมีแนวโน้มในการการประมาณค่าต่ำกว่าความเป็นจริง อย่างไรก็ตามผลการวิจัยพบว่าทั้งสองวิธีมีค่าความสัมพันธ์ระหว่างคะแนนที่แปลงได้กับคะแนนเกณฑ์ไม่ต่างกันไม่ว่าจะศึกษาในโครงสร้างลักษณะใดหรือความสัมพันธ์ระหว่างมิติความสามารถอยู่ในระดับใด

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

อภิปรายผลการวิจัย

การอภิปรายผลในงานวิจัยนี้นำเสนอ 2 ประเด็นหลัก คือ การอภิปรายผลตามสมมติฐานการวิจัย และการอภิปรายผลจากการจำลองข้อมูล มีรายละเอียดดังต่อไปนี้

1. การอภิปรายผลตามสมมติฐานการวิจัย

1.1 การอภิปรายตามสมมติฐานการวิจัยข้อที่ 1 การศึกษาปฏิสัมพันธ์ระหว่างวิธีการเชื่อมโยงคะแนนกับเงื่อนไขที่ศึกษา

จากผลการวิจัย พบว่า มีปฏิสัมพันธ์ระหว่างวิธีการเชื่อมโยงคะแนนและเงื่อนไขที่ทำการศึกษา คือ โครงสร้างมิติความสามารถ และระดับความสัมพันธ์ของมิติความสามารถ ที่ส่งผลต่อคุณภาพของการเชื่อมโยงคะแนน นั่นคือมีปฏิสัมพันธ์อย่างมีนัยสำคัญ ซึ่งพบว่าสอดคล้องกับสมติฐานการวิจัย และเป็นการขยายข้อค้นพบจากการศึกษาของ Min (2003) และ Yon (2006) ที่ได้ทำการศึกษาปฏิสัมพันธ์ระหว่างวิธีการเชื่อมโยงคะแนนกับเงื่อนไขของการศึกษา โดยผลการศึกษาครั้งนี้มีค้นพบเพิ่มเติมว่า การศึกษาปฏิสัมพันธ์ 3 ทาง จากระดับที่ 1 ไป 2 ต่างจากระดับที่ 2 ไป 3 นั่นคือ ผลการศึกษาปฏิสัมพันธ์ของวิธีการเชื่อมโยงคะแนน (METHOD) โครงสร้างมิติความสามารถ (DS) และระดับความสัมพันธ์ระหว่างมิติความสามารถ (r) เมื่อพิจารณาผลการแปลงค่าพารามิเตอร์จากระดับที่ 1 ไปยังระดับที่ 2 พบว่า METHOD X DSX r มีปฏิสัมพันธ์ที่ส่งผลต่อค่าความสัมพันธ์ระหว่างคะแนนที่แปลงได้กับคะแนนเกณฑ์ (CORR) ต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่ไม่ส่งผลต่อความคงที่และความถูกต้องของการเชื่อมโยงคะแนน

ในขณะที่ผลการแปลงคะแนนจากระดับที่ 2 ไปยังระดับที่ 3 พบว่า METHOD X DSX r มีปฏิสัมพันธ์ซึ่งส่งผลต่อความถูกต้องของการเชื่อมโยงคะแนน (BIAS) ที่ต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่ไม่ส่งผลต่อความคงที่ (RMSE) และค่าความสัมพันธ์ระหว่างคะแนนที่แปลงได้กับคะแนนเกณฑ์ (CORR) ของการเชื่อมโยงคะแนน

1.2 การอภิปรายตามสมมติฐานการวิจัยข้อที่ 2 การเปรียบเทียบคุณภาพของการเชื่อมโยงคะแนน

จากการเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนนตามโมเดล MIRT ที่มีลักษณะการหมุนแกนแตกต่างกัน 2 วิธี คือ วิธีการหมุนแกนแบบตั้งฉากโดยวิธี M (Min's Method) และวิธีการหมุนแกนแบบไม่ตั้งฉากโดยวิธี NOP (Non-Orthogonal Procrustes Method) ภายใต้โครงสร้างมิติความสามารถและระดับความสัมพันธ์ระหว่างมิติความสามารถที่ต่างกัน พบว่ามีทั้งส่วนสอดคล้องและไม่สอดคล้องกับสมมติฐานการวิจัยข้อที่ 2 ในส่วนที่สอดคล้องกับสมมติฐานการวิจัย พบว่า คุณภาพการเชื่อมโยงคะแนนตามวิธีการหมุนแกนแบบตั้งฉากโดยวิธี M

เหมาะสมสำหรับโครงสร้างที่ไม่ซับซ้อน (APSS) และมิติความสามารถที่ไม่มีความสัมพันธ์ระหว่างกัน (ρ =0) ซึ่งข้อค้นพบที่ได้เป็นการสนับสนุนผลการศึกษาของ Min (2003) และวิธี NOP เหมาะสมกับโครงสร้างที่ซับซ้อน (Mixed Structure; MS) กับระดับสูง (ρ =0.8) ซึ่งข้อค้นพบที่ได้เป็นการสนับสนุนผลการศึกษาของ Reckase & Martineau (2004) ที่นำหลักการของ Mulaik (1972) มาช่วยในการแปลงค่าพารามิเตอร์ของข้อสอบ โดยผู้สอบแต่ละคน สามารถมีค่าความสามารถในมิติที่ 1 (θ_1) และ มิติที่ 2 (θ_2) สัมพันธ์กันได้ เพื่อให้สอดคล้องกับสถานการณ์จริงเพิ่มขึ้น และเพิ่มความยืดหยุ่นในการศึกษาโดยใช้โมเดล MIRT

ในส่วนที่ไม่สอดคล้องกับสมมติฐานการวิจัย คือ ผลการวิจัยพบว่าทั้งสองวิธีมีคุณภาพของการเชื่อมโยงคะแนนไม่ต่างกันสำหรับการศึกษากับชุดของข้อสอบที่มีโครงสร้างไม่ซับซ้อน โดยผู้สอบมีความสัมพันธ์ระหว่างมิติความสามารถในระดับ 0.0 ขัดแย้งกับสมมติฐานการวิจัยที่คาดว่าวิธี M น่าจะดีกว่าวิธี NOP อาจเนื่องมาจากวิธี NOP เป็นวิธีการที่มีความยืดหยุ่นโดยอาศัยหลักการของการแปลงคะแนนแบบไม่ตั้งฉาก ประกอบกับการพัฒนาวิธีการเชื่อมโยงคะแนนโดยวิธี NOP บางส่วนมีการพัฒนาจุดอ่อนของวิธี M ซึ่งใช้หลักการทางคณิตศาสตร์เข้ามาช่วย ทั้งสองวิธีจึงมีค่าที่ใกล้เคียงกันเมื่อศึกษาในเงื่อนไขดังกล่าว

เมื่อพิจารณาระดับความสัมพันธ์ของมิติความสามารถในการเปรียบเทียบระหว่างวิธี M และวิธี NOP ของค่าพารามิเตอร์อำนาจจำแนกในมิติที่ 1 และ 2 พบว่า เมื่อไม่มีความสัมพันธ์ระหว่างมิติ วิธี M มีคุณภาพของการเชื่อมโยงคะแนนไม่แตกต่างจากวิธี NOP อย่างเห็นได้ชัด แสดงว่าในสถานการณ์ที่ไม่มีความสัมพันธ์ระหว่างมิติความสามารถสามารถเลือกใช้วิธีใดก็ได้ให้ผลไม่แตกต่างกัน ซึ่งสาเหตุสำคัญอาจเนื่องมาจากผลที่ได้จากการแปลงค่าพารามิเตอร์ของแบบสอบปรับเทียบคะแนนไปยังแบบสอบฐาน (CORR) ของทั้งสองวิธีให้ค่าที่ไม่ต่างกันมาก และทั้งสองวิธีที่ค่าที่ใช้เป็นเกณฑ์เดียวกัน ดังนั้น CORR ของทั้งสองวิธีใกล้เคียงกัน อย่างไรก็ตามเมื่อความสัมพันธ์ระหว่างมิติเพิ่มสูงขึ้น โดยเฉพาะค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.8 ควรเลือกใช้วิธี NOP สาเหตุสำคัญอาจเนื่องมาจากในระดับความสัมพันธ์ดังกล่าวพบว่า ค่า CORR ทั้งสองวิธีแตกต่างกันอย่างชัดเจน โดยวิธี NOP มีค่าใกล้เคียงกับคะแนนที่เป็นเกณฑ์มากกว่าวิธี M อย่างเห็นได้ชัด จึงส่งผลทำให้วิธี NOP มีค่า CORR สูงกว่าวิธี M ตามไปด้วย อย่างไรก็ตามเมื่อความสัมพันธ์ระหว่างมิติเพิ่มขึ้นค่าสัมประสิทธิ์สหสัมพันธ์ลดลงอยู่ในระดับปานกลาง อาจเนื่องจากเมื่อความสัมพันธ์ระหว่างมิติเพิ่มขึ้นโดยเฉพาะในโครงสร้างที่มีความซับซ้อนเกิดความคลาดเคลื่อนในการประมาณค่าสูงขึ้น โดยเฉพาะเมื่อพิจารณาจากค่า BIAS และ RMSE ดังนั้นค่าที่ได้จากการแปลงคะแนนจึงมีความสัมพันธ์กันน้อยกว่าเงื่อนไขอื่นๆ

ในขณะที่วิธีการหมุนแกนแบบไม่ตั้งฉากโดยวิธี NOP (Non-Orthogonal Procrustes Method) น่าจะ โดยพิจารณาจากดัชนีค่าความลำเอียง (BIAS) ค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) และค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) ของผลที่ได้จากการแปลงคะแนนของแบบสอบปรับเทียบคะแนนกับแบบสอบฐาน

อย่างไรก็ตามสมมติฐานดังกล่าวจะเป็นจริงเมื่อศึกษาค่าพารามิเตอร์การแปลงคะแนนค่าอำนาจจำแนกของข้อสอบในมิติที่ 1 และ มิติที่ 2 เท่านั้นจึงจะพบความแตกต่างของดัชนีบ่งชี้คุณภาพดังกล่าว แต่เมื่อศึกษาจากค่าพารามิเตอร์ความยากง่ายของข้อสอบ (d) พบว่าคุณภาพของการเชื่อมโยงคะแนนทั้งสองวิธีไม่แตกต่างกันในทุกสถานการณ์ที่ศึกษา แม้จะใช้วิธีการแปลงคะแนนของค่าพารามิเตอร์แตกต่างกัน นั่นคือ ในการแปลงคะแนนค่า d จำเป็นต้องอาศัยเมตริกซ์การหมุนแกน (Rotation Matrix; T) และเวคเตอร์การแปลงคะแนน (Translation Vector; m_1, m_2) ซึ่งสามารถแสดงสมการการแปลงคะแนนของวิธี M (สมการที่ 5.1 และ 5.2) และวิธี NOP (สมการที่ 5.3 และ 5.4) ต่อไปนี้

วิธี M $\quad a_i^* = a_i TK$ .....(5.1)

$\quad d_i^* = d_i - a_i^* Tm$ .....(5.2)

วิธี NOP $\quad a_i^* = a_i T$ .....(5.3)

$\quad d_i^* = d_i + a_i^* Tm$ .....(5.4)

เมื่อพิจารณาค่า T ของวิธี M พบว่าอาศัยหลักการหมุนแกนแบบตั้งฉาก (Orthoganal Procrustes Rotation) ตามแนวคิดของ SchÖnemann (1966) ซึ่งเป็นวิธีการที่มีความยุ่งยากและซับซ้อน โดยการวิจัยวิจัยครั้งนี้ผู้วิจัยต้องพัฒนาโปรแกรมในการคำนวณจากโปรแกรม MATLAB ดังแสดงในภาคผนวก ในขณะที่วิธี NOP ค่า T อาศัยหลักการหมุนแกนแบบไม่ตั้งฉากตามแนวคิดของ Mulaik (1972) โดยคำนวณจากเมตริกซ์ของค่าพารามิเตอร์อำนาจจำแนกของแบบสอบปรับเทียบคะแนน (A) และเมตริกซ์ของค่าพารามิเตอร์อำนาจจำแนกของแบบสอบฐาน (B) ซึ่ง Reckase & Martineau (2004) ได้นำแนวคิดดังกล่าวมาใช้ สามารถคำนวณได้โดยใช้โปรแกรม EXCEL หรือคำนวณด้วยมือได้ ดังแสดงในสมการที่ 5.5

$$T = (A'A)^{-1}A'B \quad \text{.....(5.5)}$$

จากผลการคำนวณที่ใช้หลักการที่แตกต่างกันส่งผลให้ทั้งสองวิธีมีค่า T ที่แตกต่างกัน

เมื่อพิจารณาเวคเตอร์การแปลงคะแนน (Translation Vector; m_1, m_2) ซึ่งถือเป็นองค์ประกอบที่สำคัญยิ่งของการแปลงค่าพารามิเตอร์ d พบว่าทั้งสองวิธีมีค่าที่แตกต่างกันทั้ง m1 และ m2 อย่างไรก็ตามทั้งสองวิธีอาศัยหลักการคำนวณเวคเตอร์การแปลงคะแนนเดียวกัน โดยใช้วิธีกำลังสองน้อยที่สุด (Least-squares Solution) ที่พัฒนาโดย Li & Lissitz (2000) เพื่อให้เกิดความแตกต่างระหว่างค่าพารามิเตอร์จากแบบสอบฐานกับค่าพาราเตอร์การแปลงคะแนนจากแบบสอบปรับเทียบคะแนนไปยังแบบสอบฐาน ซึ่งสามารถแสดงได้ดังสมการที่ 5.6

$$Q = \sum_{i=1}^{n} \left(d_{iB} - d_{iB}^* \right)^2 \quad \text{.....(5.6)}$$

อย่างไรก็ตามทั้งสองวิธีต่างกันตรงที่ค่าเวคเตอร์การแปลงคะแนนของวิธี M จะคำนวณตามวิธีของ Li & Lissitz (2000) ให้ความสำคัญกับการคำนวณในแต่ละมิติ ในขณะที่วิธี NOP แม้จะอาศัยหลักการเดียวกันแต่จะมีการคำนวณโดยรวมไม่ได้มุ่งเน้นในแต่ละมิติ และมีขั้นตอนในการคำนวณแตกต่างกัน จึงทำให้เวคเตอร์การแปลงคะแนนทั้งสองวิธี แตกต่างกันไปด้วย แต่เมื่อทำการแปลงค่าพารามิเตอร์ d จากแบบสอบปรับเทียบคะแนนไปยังแบบสอบฐานซึ่งมีสมการในการแปลงคะแนนต่างกันของแต่ละวิธี ดังสมการที่ 5.2 และ 5.4 พบว่าทั้งสองวิธีมีค่าที่เท่ากันทุกสถานการณ์

สาเหตุที่ให้ผลไม่แตกต่างกันแม้ว่าวิธี M จะต้องอาศัยหลักการคำนวณเมตริกซ์การหมุนแกน (T) และเวคเตอร์การแปลงคะแนน (m_1, m_2) ที่มีความซับซ้อนและสนใจศึกษาในแต่ละมิติ แต่การศึกษาในครั้งนี้เป็นการศึกษาเพียง 2 มิติ ดังนั้นจึงไม่เกิดความคลาดเคลื่อนในการคำนวณหรือความคลาดเคลื่อนมีน้อยมาก ซึ่งผลการวิจัยในส่วนนี้สนับสนุนกับแนวคิดของ Min (2003) และ Matineua & Reckase (2004) ที่พบว่า วิธี M เหมาะสมสำหรับมิติที่ทำการศึกษามีจำนวน 2-3 มิติ (Low-Dimensional Space) ดังนั้นผลที่ได้จากจึงไม่แตกต่างจากวิธี NOP โดยมีการหมุนแกนเป็นไปโดยอัตโนมัติสำหรับการแปลงคะแนนในแต่ละมิติของเมตริกซ์ปรับเทียบคะแนนกับมิติของเมตริกซ์ฐาน ขึ้นอยู่กับความสอดคล้องของมิติในเมตริกซ์ปรับเทียบคะแนนกับมิติของเมตริกซ์ฐาน ขณะเดียวกันการแปลงคะแนนแบบไม่ตั้งฉาก ช่วยลดการคำนวณน้ำหนักของการปรับเทียบคะแนนทุกครั้งของการแปลงคะแนน ในเมตริกซ์ปรับเทียบคะแนนเป็นการลดความยุ่งยากในระดับหนึ่ง ซึ่งการตัดสินใจของมิติในเมตริกซ์ปรับเทียบคะแนน คือการจับคู่กับมิติในเมตริกซ์ฐาน โดยใช้กระบวนการทางคณิตศาสตร์มาช่วยคำนวณถือได้ว่าเป็นวิธีการที่ง่ายและไม่ซับซ้อน แต่สำหรับกรณีที่โมเดลมีจำนวนมิติมากขึ้นและมีความซับซ้อนมากเกินไป (High-Dimensional Space) การคำนวณโดยวิธี M จะจัดกระทำได้ยากหรือไม่สามารถคำนวณได้ เนื่องจากปฏิสัมพันธ์ระหว่างคุณลักษณะเด่นของมิติข้ามชุดของแบบสอบ โดยเฉพาะอย่างยิ่งคุณลักษณะของการหมุนแกนแบบตั้งฉาก ผลที่ได้อาจจะเกิดความคลาดเคลื่อนในระดับที่สูงขึ้น ที่จำเป็นต้องมีการหมุนแกนการเปลี่ยนแปลงมิติของการเปรียบเทียบแต่ละคู่จากแบบสอบปรับเทียบคะแนนไปยังแบบสอบฐาน ซึ่งอาจจะส่งผลต่อค่าพารามิเตอร์การแปลงคะแนน d ได้ ซึ่งในประเด็นนี้ยังไม่มีข้อมูลเชิงประจักษ์ในการศึกษาว่าจำนวนมิติเป็นเท่าใดหรือโครงสร้างมิติความสามารถมีลักษณะเป็นอย่างไรจึงทำให้ค่าพารามิเตอร์การแปลงคะแนนของวิธี M แตกต่างจากวิธี NOP

จุฬาลงกรณ์มหาวิทยาลัย

2. การอภิปรายผลข้อมูลที่ได้จากการจำลองข้อมูล

2.1 การสร้างค่าพารามิเตอร์ที่แท้จริงของข้อสอบและค่าพารามิเตอร์ที่ได้จากการประมาณค่า (Calibration)

สำหรับขั้นตอนในการสร้างค่าพารามิเตอร์ที่แท้จริงของข้อสอบ ประกอบด้วย a_1 a_2 และ d จำเป็นต้องมีการกำหนดมุมในแต่ละมิติเพื่อนำไปคำนวณค่าพารามิเตอร์ของข้อสอบในแต่ละมิติ ซึ่งในการวิจัยครั้งนี้มีการกำหนดขาดของมุมจากการสุ่มแบบเอกภาพ (Uniform Distribution) ซึ่งผลการศึกษาพบว่าเมื่อกำหนดขนาดของมุมให้อยู่ในโครงสร้างที่ซับซ้อน (MS) ที่อยู่ใกล้มุม 25 หรือ 40 องศา ซึ่งเป็นขนาดของมุมที่เป็นรอยต่อระหว่างมิติที่ 1 และมิติที่ 2 พบว่าเมื่อนำไปประมาณค่าพารามิเตอร์ (Calibration) โดยใช้โปรแกรม TESTFACT ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบที่ประมาณได้มีความคลาดเคลื่อนสูง และข้อสอบข้อนั้นไม่สามารถระบุได้ว่าควรจัดอยู่ในมิติที่ 1 หรือมิติที่ 2 เนื่องจากมีค่าน้ำหนักใกล้เคียงกัน ประกอบกับข้อมูลที่ได้บางครั้งมีค่าที่สูงหรือต่ำกว่าความเป็นจริงอย่างเห็นได้ชัด ซึ่งมีความจำเป็นต้องตัดข้อมูลทิ้งและทำการสร้างข้อมูลใหม่

จากผลที่เกิดขึ้นดังกล่าวสามารถสรุปได้ว่าการกำหนดขนาดของมุมมีผลต่อความถูกต้องแม่นยำของจำนวนมิติที่ทำการศึกษา สาเหตุสำคัญอาจเนื่องมาจากค่าพารามิเตอร์อำนาจจำแนกของข้อสอบตามโมเดล MIRT มีลักษณะคล้ายคลึงกับค่าน้ำหนักองค์ประกอบ (Factor Loading) ตามโมเดลการวิเคราะห์องค์ประกอบ (Factor Analysis) ซึ่งข้อใดมีค่าอำนาจจำแนกสูงในมิติใดมิติหนึ่ง ถือว่ามีค่าน้ำหนักองค์ประกอบสูงในมิตินั้น เช่น ข้อสอบมีอำนาจจำแนกสูงในมิติที่ 1 มากกว่ามิติที่ 2 แสดงว่าข้อนั้นจะมีค่าน้ำหนักองค์ประกอบสูงในมิติที่ 1 ด้วย จึงจัดให้ข้อนั้นเหมาะสมที่จะอยู่ในมิติที่ 1 มากกว่า เป็นต้น ซึ่งสอดคล้องกับการศึกษาของ Rotou, Headrick & Elmore (2000, site in Reckase, 2006) และการศึกษาของ Ackerman & Walker (2003) ได้ระบุว่า “การประเมินว่าข้อสอบนั้นอยู่ในมิติใดสามารถพิจารณาได้จากค่าอำนาจจำแนกของข้อสอบทั้งหมดในข้อนั้น โดยข้อสอบข้อนั้นจะจัดอยู่ในมิติที่มีค่าอำนาจจำแนกมากที่สุด นั่นคือถ้ามีค่าอำนาจจำแนกในมิติแรกมากกว่ามิติอื่นๆ แสดงว่าข้อสอบข้อนั้นอยู่ในมิติแรกมากกว่ามิติอื่นๆ รวมถึงสะท้อนให้เห็นถึงตำแหน่งเวคเตอร์ข้อสอบที่ควรอยู่ใกล้กับมิติแรกมากกว่ามิติอื่นๆ ด้วยเช่นกัน” ดังนั้นเพื่อให้การประมาณค่าพารามิเตอร์อำนาจจำแนกที่เกี่ยวข้องกับตำแหน่งของข้อสอบ (Location) เป็นไปตามเงื่อนไขที่วางไว้จึงต้องพิจารณามุมให้เหมาะสมและไม่ควรอยู่ในช่วงรอยต่อระหว่างมิติที่ 1 และมิติที่ 2 นอกจากนี้ผลการวิจัยพบว่าเมื่อการกำหนดให้มุมมีขนาด 0 หรือ 90 องศา พบว่าเมื่อนำไปประมาณค่าโดยใช้โปรแกรม TESTFACT พบว่ามีอำนาจจาจจำแนกติดลบ ซึ่งมีแนวโน้มในการประมาณค่าที่ต่ำกว่าความเป็นจริง

สำหรับการกำหนดค่าพารามิเตอร์โอกาสการเดาของข้อสอบ (C) เพื่อนำไปใช้ในการคำนวณ $P(U_{ij}=1|a_i,d_i,\theta_j,c_i)$ ซึ่งเป็นความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง (ได้ 1 คะแนน) สำหรับข้อสอบข้อที่ i สำหรับบุคคลที่ j ในมิติที่ m พบว่ามีส่วนสำคัญต่อความคลาดเคลื่อนของ

การประมาณค่าโอกาสการเดาของข้อสอบ (C) เนื่องจากมีการประมาณค่าที่สูงกว่าความเป็นจริง เนื่องจากผู้วิจัยได้ทำการศึกษาเบื้องต้นโดยกำหนดให้พารามิเตอร์ C ที่แท้จริงมีค่าเท่ากับ 0.25 ในทุกข้อ เมื่อนำรูปแบบการตอบข้อสอบแบบสองค่าไปประมาณค่าโดยใช้โปรแกรม BILOG พบว่ามีการประมาณค่าสูงกว่าความเป็นจริงอยู่ในช่วง 0.3-0.5 เมื่อผู้วิจัยทำการตรวจสอบค่าความเที่ยงของข้อสอบพบว่ามีค่าความเที่ยงต่ำ เนื่องจากโอกาสการเดาข้อสอบมีผลต่อค่าความเที่ยงของข้อสอบ เมื่อโอกาสการเดาเพิ่มขึ้นมีผลทำให้ค่าความเที่ยงลดลง (ศิริชัย กาญจนวาสี, 2550) นอกจากนี้ผลการศึกษาดังกล่าวเป็นการสนับสนุนผลการวิจัยของ Rogers (1999) & Yeh (2007) ที่พบว่าค่าโอกาสการเดาของข้อสอบส่งผลกระทบต่อความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง และส่งผลกระทบต่อความสัมพันธ์ระหว่างของข้อสอบในชุดของแบบสอบ นอกจากนี้ยังสอดคล้องกับการศึกษาของ Messick (1995, site in Yeh, 2007) ที่พบว่าค่าโอกาสการเดาของข้อสอบส่งผลกระทบต่อความเป็นมิติของแบบสอบ โดยเฉพาะความตรงของโครงสร้างที่ต้องการศึกษา

ดังนั้นในการศึกษาครั้งนี้ผู้วิจัยจึงมีการกำหนดให้ค่า C ได้จากการการสุ่มแบบ β (2,50) เนื่องจากให้ค่า c อยู่ในช่วงที่ยอมรับได้คือไม่เกิน 0.2 (Harwell, Stone, Hsu, & Kirisci, 1996) เมื่อนำไปประมาณค่าโดยใช้โปรแกรม BILOG-MG พบว่ามีค่าสูงกว่าความเป็นจริงในช่วงที่ยอมรับได้ อย่างไรก็ตามในการกำหนดค่า C สำหรับการจำลองข้อมูลเพื่อใช้ศึกษาในโมเดล MIRT ยังเป็นประเด็นที่ขาดความชัดเจนว่าควรเป็นอย่างไรจึงจะเหมาะสม และให้เกิดความคลาดเคลื่อนในการประมาณค่าน้อยที่สุด เพราะในการประมาณค่าทุกครั้งย่อมเกิดความคลาดเคลื่อนเสมอ และมีแนวโน้มที่จะสูงกว่าความเป็นจริง

2.2 การตรวจสอบค่าความเที่ยงของข้อมูลที่จำลองได้

สำหรับผลที่ได้จากการจำลองในขั้นตอนของการสร้างข้อมูล (Generate Data) พบว่าจำนวนข้อสอบร่วมเป็นปัจจัยสำคัญที่ส่งผลต่อค่าความเที่ยงของข้อสอบโดยเฉพาะการศึกษาที่มีหลายมิติ เนื่องจากผู้วิจัยได้ทำการศึกษาเบื้องต้น (Pilot Study) กับจำนวนข้อสอบร่วมเพียง 20 ข้อ พบว่าโครงสร้างมิติความสามารถและความสัมพันธ์ระหว่างมิติความสามารถเป็นไปตามเงื่อนไขที่ศึกษา แต่มีค่าความเที่ยงของข้อสอบต่ำโดยเฉพาะในเงื่อนไขที่ไม่มีความสัมพันธ์ระหว่างมิติความสามารถ (r= 0.0) ซึ่งมีความเที่ยงอยู่ในช่วง 0.40-0.55 แต่เมื่อความสัมพันธ์ระหว่างมิติสูงขึ้นค่าความเที่ยงมีค่าเพิ่มขึ้นแต่มีค่าไม่เกิน 0.70 สาเหตุสำคัญคือข้อสอบมีน้อยเกินไปสำหรับใช้ในการศึกษา 2 มิติ ในกลุ่มตัวอย่างขนาดใหญ่ โดยผลการศึกษาดังกล่าวสอดคล้องกับการศึกษาของ Li & Lissize (2000) และ Reckase & Matineua (2000) ที่พบว่าจำนวนข้อสอบร่วมเป็นสิ่งที่สำคัญ จากผลการศึกษาดังกล่าวพบว่าจำนวนข้อสอบร่วมที่ยาวกว่าย่อมให้ผลที่ดีกว่า ซึ่งจากการศึกษาของ Li & Lissitz (2000) ได้ศึกษาจำนวนของข้อสอบร่วม 15 ข้อ และ 25 ข้อ และ Reckase & Martineau (2004, 2006) ได้ใช้

ข้อสอบร่วม ระดับละ 25 ข้อ ข้อค้นพบจากการศึกษาดังกล่าวพบว่าข้อสอบร่วมภายในน้อยเกินไปกับกลุ่มผู้สอบที่มีขนาดใหญ่

ดังนั้นการวิจัยในครั้งนี้ผู้วิจัยจึงเพิ่มจำนวนข้อสอบร่วมจาก 20 ข้อ เป็น 30 ข้อ เมื่อทำการตรวจสอบค่าความเที่ยงของข้อสอบพบว่ามีค่าความเที่ยงเพิ่มขึ้นอย่างชัดเจน โดยข้อมูลที่จำลองได้ในเงื่อนไขที่ไม่มีความสัมพันธ์ระหว่างมิติความสามารถ พบว่ามีค่าความเที่ยงเพิ่มขึ้นอย่างชัดเจนตั้งแต่ 0.65-0.75 ซึ่งโครงสร้างที่ซับซ้อน (MS) จะมีความเที่ยงสูงกว่าโครงสร้างที่ไม่ซับซ้อน (APSS) ที่มีค่าสัมประสิทธิ์สหสัมพันธ์ในระดับเดียวกัน เมื่อทำการตรวจสอบค่าความเที่ยงในเงื่อนไขที่ความสัมพันธ์ระหว่างมิติความสามารถอยู่ในระดับปานกลางและระดับสูง พบว่ามีค่าความเที่ยงสูงอยู่ในช่วงตั้งแต่ 0.71-0.92 ดังนั้นผู้วิจัยจึงเลือกศึกษาโดยใช้ข้อสอบร่วม 30 ข้อ เพื่อเพิ่มความน่าเชื่อถือของข้อมูลที่ได้และให้เป็นไปตามเงื่อนไขที่ศึกษา อย่างไรก็ตามประเด็นในเรื่องการใช้ข้อสอบร่วมจำนวนเท่าใดจึงจะเพียงพอต่อการเชื่อมโยงคะแนนตามโมเดล MIRT ยังไม่มีความชัดเจนในประเด็นนี้

2.3 เกณฑ์ที่ใช้ในการเปรียบเทียบคุณภาพของการเชื่อมโยงคะแนน

เกณฑ์ที่ใช้ในการวิจัยครั้งนี้พิจารณาจากดัชนี RMSE ที่สะท้อนถึงความคงที่ของการเชื่อมโยงคะแนน BIAS ที่สะท้อนถึงความถูกต้องของการเชื่อมโยงคะแนน และ CORR ที่สะท้อนถึงความสัมพันธ์ระหว่างคะแนนที่แปลงได้กับคะแนนเกณฑ์ ซึ่งจากผลการวิจัยพบว่าเกณฑ์ที่เหมาะสมในการนำมาใช้ในการเปรียบเทียบคุณภาพของการเชื่อมโยงคะแนนมี 2 ดัชนี คือ RMSE และ BIAS เนื่องจากแปรผันไปตามเงื่อนไขที่ศึกษาซึ่งสะท้อนให้เห็นถึงความคงที่และความถูกต้องของการเชื่อมโยงคะแนนต่างกันระหว่างวิธี M และ วิธี NOP แต่เมื่อใช้เกณฑ์ CORR พบว่า วิธีทั้งสองมีค่าที่ใกล้เคียงกันและไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ซึ่งไม่สามารถสะท้อนให้เห็นถึงความแตกต่างของคะแนนที่แปลงได้กับคะแนนเกณฑ์ได้อย่างชัดเจน

สาเหตุสำคัญอาจเนื่องมาจากผลที่ได้จากการแปลงค่าพารามิเตอร์ของแบบสอบปรับเทียบคะแนนไปยังแบบสอบฐาน (CORR) ของทั้งสองวิธีให้ค่าที่ไม่ต่างกันมาก และทั้งสองวิธีที่ค่าที่ใช้เป็นเกณฑ์เดียวกัน ดังนั้น CORR ของทั้งสองวิธีใกล้เคียงกัน อย่างไรก็ตามเมื่อความสัมพันธ์ระหว่างมิติเพิ่มสูงขึ้น โดยเฉพาะค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.8 ควรเลือกใช้วิธี NOP สาเหตุสำคัญอาจเนื่องมาจากในระดับความสัมพันธ์ดังกล่าวพบว่า ค่า CORR ทั้งสองวิธีแตกต่างกันอย่างชัดเจน โดยวิธี NOP มีค่าใกล้เคียงกับคะแนนที่เป็นเกณฑ์มากกว่าวิธี M อย่างเห็นได้ชัด จึงส่งผลทำให้วิธี NOP มีค่า CORR สูงกว่าวิธี M ตามไปด้วย อย่างไรก็ตามเมื่อความสัมพันธ์ระหว่างมิติเพิ่มขึ้นค่าสัมประสิทธิ์สหสัมพันธ์ลดลงอยู่ในระดับปานกลาง อาจเนื่องจากเมื่อความสัมพันธ์ระหว่างมิติเพิ่มขึ้นโดยเฉพาะในโครงสร้างที่มีความซับซ้อนเกิดความคลาดเคลื่อนในการประมาณค่าสูงขึ้นโดยเฉพาะเมื่อพิจารณาจากค่า BIAS และ RMSE ดังนั้นค่าที่ได้จากการแปลงคะแนนจึงมีความสัมพันธ์กันน้อยกว่าเงื่อนไขอื่นๆ

ข้อเสนอแนะ

1. ข้อเสนอแนะสำหรับการนำไปใช้

ในที่นี้ผู้วิจัยมีข้อเสนอแนะสำหรับการนำไปใช้ในประเด็นของการเลือกใช้ในสถานการณ์จริง และข้อเสนอแนะในเชิงเทคนิค โดยมีรายละเอียดดังนี้

1.1 ข้อเสนอแนะในการนำไปประยุกต์ใช้ในสถานการณ์จริง

1.1.1 การวิจัยในครั้งนี้พบว่า วิธี M และวิธี NOP เป็นวิธีการเชื่อมโยงคะแนนที่มีความถูกต้อง (Accuracy) และมีความคงที่ (Stable) ใกล้เคียงกันเมื่อศึกษาในเงื่อนไขที่ไม่มีความสัมพันธ์ระหว่างมิติ ดังนั้นเมื่อต้องการศึกษาในเงื่อนไขดังกล่าวสามารถเลือกใช้วิธีใดวิธีหนึ่งได้ เพราะให้ผลไม่แตกต่างกันมากนัก แต่เมื่อความสัมพันธ์ระหว่างมิติเพิ่มสูงขึ้น พบว่าวิธี NOP มีความถูกต้องและคงที่มากกว่า โดยเฉพาะในโครงสร้างที่มีความซับซ้อนมากขึ้น ดังนั้นควรเลือกใช้วิธี NOP เนื่องจากเทคนิคที่ใช้มีความเหมาะสมและสอดคล้องกับข้อมูลมากกว่าวิธี M

จากผลดังกล่าวสามารถเป็นแนวทางในการนำไปใช้วัดผลสัมฤทธิ์ทางการเรียนที่มีลักษณะการบูรณาการความรู้ข้ามเนื้อหาได้ทุกระดับช่วงชั้น และทุกลุ่มสาระการเรียนรู้ เพื่อวัดพัฒนาการการเรียนรู้ของผู้เรียนจากระดับชั้นหนึ่งไปอีกระดับชั้นหนึ่งที่มีลักษณะภาคตัดขวาง (Cross-sectional) ซึ่งเป็นการศึกษาในภาคการศึกษาหรือปีการศึกษาเดียวกัน ซึ่งสิ่งสำคัญในการนำไปใช้คือต้องมีการตรวจสอบว่าแบบสอบที่ใช้มีโครงสร้างเป็นไปตามเงื่อนไขที่กำหนดไว้หรือไม่ โดยเฉพาะกลุ่มสาระการเรียนรู้คณิตศาสตร์และวิทยาศาสตร์ที่มีกลุ่มของเนื้อหาค่อนข้างเด่นชัด

1.1.2 ผลการวิจัยพบว่าค่าพารามิเตอร์ความยากของข้อสอบ (d) ที่ได้จากการแปลงคะแนนจากแบบสอบปรับเทียบคะแนนไปยังแบบสอบฐานมีค่าเท่ากันทั้งสองวิธี สำหรับการศึกษาใน 2 มิติ โดยทั้งสองวิธีจำเป็นต้องอาศัยเมตริกซ์การหมุนแกนและเวคเตอร์การแปลงคะแนนเหมือนกัน แต่มีหลักการในการคำนวณแตกต่างกันอย่างชัดเจน โดยการใช้วิธี NOP ในการเชื่อมโยงคะแนนถือได้ว่าเป็นวิธีการที่ง่ายและไม่ซับซ้อน ดังนั้นวิธี NOP จึงสะดวกต่อการนำไปประยุกต์ใช้มากกว่าวิธี M ในแปลงค่าความยากง่ายของข้อสอบเนื่องจากให้คุณภาพของการเชื่อมโยงคะแนนไม่แตกต่างกัน

1.1.3 เนื่องจากการจำลองข้อมูลในครั้งนี้ กำหนดขึ้นเพื่อให้สามารถนำไปประยุกต์ใช้ได้หลากหลายสถานการณ์ โดยเฉพาะการศึกษากับการวัดความสามารถผู้สอบใน 2 มิติ โดยกลุ่มผู้สอบมีระดับความสามารถต่างกัน ทั้งในด้านการวัดความสามารถและการวัดผลสัมฤทธิ์ทางการเรียน สำหรับการวัดความสามารถของผู้สอบ สามารถนำไปใช้เพื่อเชื่อมโยง

คะแนนที่มีลักษณะ 2 มิติ เช่นการวัดความสามารถทางด้านภาษาและด้านการคำนวณ เป็นต้น ส่วนการนำไปใช้วัดผลสัมฤทธิ์ทางการเรียนทีมีลักษณะการบูรณาการความรู้ข้ามเนื้อหาสามารถนำไปประยุกต์ใช้ได้ทุกระดับช่วงชั้น และทุกลุ่มสาระการเรียนรู้เพื่อวัดพัฒนาการการเรียนรู้ของผู้เรียนจากระดับชั้นหนึ่งไปอีกระดับชั้นหนึ่งที่มีลักษณะภาคตัดขวาง (Cross-Sectional) ซึ่งเป็นการศึกษาในภาคการศึกษาหรือปีการศึกษาเดียวกัน โดยเฉพาะกลุ่มสาระการเรียนรู้คณิตศาสตร์และวิทยาศาสตร์ที่มีกลุ่มของเนื้อหาค่อนข้างเด่นชัด เช่น ในกลุ่มสาระคณิตศาสตร์ประกอบด้วย 5 กลุ่มที่สำคัญ คือ 1) จำนวนและการดำเนินการ 2) การวัด 3) เรขาคณิต 4) พีชคณิต และ 5) การวิเคราะห์ข้อมูลและความน่าจะเป็น สำหรับกรณีที่ศึกษาในรายวิชาวิทยาศาสตร์สามารถแยกเป็นกลุ่มเนื้อหาที่สำคัญ คือ วิทยาศาสตร์ชีววิทยา (Biology Science) วิทยาศาสตร์กายภาพ (Physical Science) และวิทยาศาสตร์สิ่งแวดล้อม (Environmental Science) เป็นต้น

จากที่กล่าวมาสามารถศึกษาได้คราวละ 2 มิติ โดยทำการตรวจสอบความเป็นพหุมิติจากหลักฐานทางสถิติโดยการวิเคราะห์ตามโมเดล MIRT ร่วมกับการกำหนดโครงสร้างเนื้อหาของแบบสอบ (Test Specification) เพื่อเป็นแนวทางในการเขียนข้อสอบตามเนื้อหาและหน่วยที่ต้องการวัดที่แสดงถึงความเป็นตัวแทนของมิติที่มุ่งวัดได้อย่างถูกต้องและการวิเคราะห์เนื้อหา (Content Analysis) เพื่อเป็นแนวทางในการตรวจสอบมิติ ซึ่งมี 2 วิธี วิธีแรกอาจให้ผู้เชี่ยวชาญที่มีประสบการณ์ในการตรวจสอบหรือตัดสินมิติที่ต้องการวัด หรือวิธีที่สองการตัดสินโดยการยึดเนื้อหาเป็นหลัก อย่างไรก็ตามการศึกษาการเชื่อมโยงคะแนนแบบพหุมิติเหมาะสำหรับการสอบที่มีกลุ่มตัวอย่างขนาดใหญ่ (Large Scale) โดยในที่นี้ผู้วิจัยได้เสนอตัวอย่างของขั้นตอนในการนำไปประยุกต์ใช้กับการเชื่อมโยงคะแนนที่มีลักษณะ 2 มิติ ดังนี้

1) เลือกวิธีการเก็บรวบรวมข้อมูล และเก็บรวบรวมข้อมูลตามรูปแบบที่กำหนดไว้ ซึ่งในที่นี้ได้ยกตัวอย่างการออกแบบการเก็บรวบรวมข้อมูลโดยใช้รูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้แบบสอบร่วม (Anchor-Test Nonequivalent Group Design) เนื่องจากเป็นรูปแบบที่ง่ายต่อการนำไปปฏิบัติมากกว่ารูปแบบอื่น โดยการสร้างข้อสอบร่วม (Common Item) ประมาณร้อยละ 30 ของจำนวนข้อสอบทั้งฉบับ โดยนักเรียนในแต่ละระดับจะทำแบบสอบในระดับตัวเอง และระดับชั้นที่สูงกว่าหรือต่ำกว่า ซึ่งถือเป็นข้อสอบร่วม แนวทางในการสร้างข้อสอบร่วมคือให้มีความเป็นตัวแทนที่เพียงพอทั้งในมิติที่ 1 และมิติที่ 2 โดยการพิจารณาจากโครงสร้างเนื้อหา (Test Specification) ซึ่งเป็นตัวสะท้อนที่สำคัญว่าชุดของข้อสอบมีโครงสร้างมิติลักษณะใดเป็นโครงสร้างที่ซับซ้อน (MS) หรือไม่ซับซ้อน (APSS)

2) การประมาณค่าพารามิเตอร์ (Calibration) แบบสอบแต่ละชุด โดยใช้โปรแกรม BILOG-MG ในการประมาณค่าพารามิเตอร์โอกาสการเดาของข้อสอบ (c) และนำค่าพารามิเตอร์ที่ได้ประมาณค่าพารามิเตอร์ความยากข้อสอบ (d) ค่าอำนาจจำแนกของข้อสอบในมิติที่ 1 และมิติที่ 2 รวมถึงค่าความสามารถของผู้สอบโดยใช้โปรแกรม TESTFACT ซึ่ง

ค่าพารามิเตอร์ของแบบสอบในแต่ละคู่จะถูกหมุนแกน (Rotated) และ ถูกแปลงคะแนนให้อยู่ในสเกลเดียวกัน (Common Configuration) โดยเลือกใช้กระบวนการหมุนแกนแบบตั้งฉาก (Orthogonal Procrustes Rotation) หรือการหมุนแกนแบบไม่ตั้งฉาก (Non-orthogonal Procrustes Rotation) แล้วแต่สถานการณ์ที่ต้องการศึกษา

3) แปลงค่าพารามิเตอร์ของข้อสอบจากแบบสอบระดับชั้นหนึ่งไปยังระดับชั้นหนึ่ง โดยพิจารณาเลือกใช้วิธีการเชื่อมโยงคะแนนให้เหมาะสม สำหรับผลการวิจัยในครั้งนี้ควรเลือกใช้วิธี NOP เนื่องจากเทคนิคที่ใช้มีความเหมาะสมและสอดคล้องกับข้อมูลมากกว่าวิธี M และสะดวกต่อการนำไปใช้มากกว่า โดยมีการแปลงทั้งค่าพารามิเตอร์ของข้อสอบโดยใช้โปรแกรมที่ผู้วิจัยพัฒนาขึ้นดังแสดงในภาคผนวก ง ส่วนการแปลงค่าพารามิเตอร์ความสามารถของผู้สอบ (θ) ใช้สมการการแปลงคะแนน ดังนี้

วิธี NOP $\theta_i^* = T^{-1}\theta_j - m$ .....(5.7)

เมื่อ T (Rotation Matrix) แทน เมตริกซ์การหมุนแกน และ m (Translation Vector) แทน เวคเตอร์การแปลงคะแนน ซึ่งสามารถนำค่า T และ m จากโปรแกรมที่ผู้วิจัยพัฒนาขึ้นมาใช้ในการแปลงค่าพารามิเตอร์ของผู้สอบได้

4) สร้างตารางการแปลงคะแนนค่าพารามิเตอร์ของข้อสอบร่วม และค่าความสามารถที่แท้จริงข้อผู้สอบซึ่งโดยส่วนใหญ่จะกำหนดให้อยู่ในช่วง -3 ถึง +3 ภายหลังจากการแปลงคะแนนค่าพารามิเตอร์ของข้อสอบให้มีโครงสร้างร่วมกัน (Coordinate Framework) ในสเปซแบบพหุมิติ (Multidimensional Space)

5) ประเมินผลการเชื่อมโยงคะแนน โดยเลือกใช้เกณฑ์การประเมินตามความเหมาะสม โดยเฉพาะการพิจารณาความคงที่ของการเชื่อมโยงคะแนนจากดัชนี RMSE จะเหมาะสมก็ต่อเมื่อเมื่อมีค่าน้อย รวมถึงการพิจารณาค่าความสัมพันธ์ระหว่างคะแนนที่แปลงได้กับคะแนนเกณฑ์ (CORR) จะเหมาะสมก็ต่อเมื่อมีค่าความสัมพันธ์ในระดับสูง แต่อย่างไรก็ตามเกณฑ์การพิจารณาจากค่า CORR ส่วนการพิจารณาความถูกต้องของการเชื่อมโยงคะแนนจากดัชนี BIAS จะใช้ได้เฉพาะกรณีที่ทราบค่าพารามิเตอร์ที่แท้จริงเท่านั้น

6) แปลความหมายของคะแนนที่ได้เพื่อศึกษาการเปลี่ยนแปลงในแต่ละมิติแฝง และศึกษาจากคะแนนรวม (Composite Scores) โดยพิจารณาความแตกต่างของคะแนนในตำแหน่งที่ต่างกันของคะแนนที่ปรับให้อยู่ในสเกลเดียวกัน ซึ่งเวคเตอร์ข้อสอบสามารถนำมาพล็อตกราฟเพื่อเปรียบเทียบทักษะและความรู้เพื่อนำมาใช้ในการศึกษาพัฒนาการในแต่ละระดับความสามารถที่ต่างกัน

1.2 ข้อเสนอแนะในเชิงเทคนิคของการจำลองข้อมูล

1.2.1 ในขั้นตอนของการจำลองข้อมูลพบว่าการกำหนดขนาดของมุมเพื่อสร้างสถานการณ์ในการจำลองข้อมูล ให้มีจำนวน 2 มิติ ซึ่งมีโครงสร้างมิติความสามารถและความสัมพันธ์ระหว่างมิติความสามารถที่แตกต่างกัน ต้องมีการกำหนดขนาดของมุมให้เหมาะสม โดยเฉพาะอย่างยิ่งการสุ่มมุมที่อยู่ใกล้รอยต่อระหว่างมิติที่ 1 และ มิติที่ 2 แม้ว่าจะได้ค่าพารามิเตอร์ที่แท้จริงเป็นไปตามเงื่อนไขที่กำหนดไว้ แต่เมื่อนำไปประมาณค่าพารามิเตอร์โดยโปรแกรม TESTFACT มักเกิดความคลาดเคลื่อนสูงเสมอ โดยเฉพาะอย่างยิ่งค่าอำนาจจำแนกของข้อสอบซึ่งมีแนวโน้มที่จะสูงกว่าความเป็นจริง ดังนั้นเมื่อได้รูปแบบการตอบข้อสอบที่คำนวณได้จากค่าพารามิเตอร์ที่แท้จริงแล้ว ผู้วิจัยควรทำการตรวจสอบว่าเป็นไปตามเงื่อนไขที่กำหนดไว้หรือไม่ เพื่อให้ข้อมูลมีความถูกต้องก่อนที่จะนำไปสู่ขั้นตอนของการเปรียบเทียบคุณภาพของการเชื่อมโยงคะแนนต่อไป สำหรับการประมาณค่าพาริเตอร์ความยากของข้อสอบ (d) พบว่ามักมีค่าต่ำกว่าความเป็นจริงในบางข้อ ดังนั้นในการสร้างค่าค่วามยากของข้อสอบที่เป็นค่าที่แท้จริงควรอยู่ในช่วงที่เหมาะสม ไม่ควรมีค่าสูงหรือต่ำจนเกินไป ซึ่งจากผลการศึกษาควรมีค่าไม่เกิน ± 0.20

1.2.2 ในขั้นตอนของการกำหนดขนาดของข้อสอบร่วม ผู้วิจัยพบว่าการใช้จำนวนข้อสอบร่วมภายในร้อยละ 20 ของจำนวนข้อสอบทั้งหมดไม่เหมาะสมสำหรับการศึกษาใน 2 มิติ ดังนั้นการกำหนดขนาดจำนวนข้อสอบร่วมเป็นสิ่งสำคัญในการศึกษาโดยใช้โมเดล MIRT ซึ่งจำนวนข้อสอบควรมีความเป็นตัวแทนที่เพียงพอของแต่ละมิติที่ศึกษา ซึ่งถ้าข้อสอบมีจำนวนไม่เพียงพอ จะส่งผลต่อค่าความเที่ยงของข้อสอบอย่างเห็นได้ชัด อย่างไรก็ตามในการศึกษาครั้งนี้ผู้วิจัยใช้ข้อสอบร่วมภายในประมาณร้อยละ 30 ของข้อสอบทั้งหมด ซึ่งตรวจสอบพบว่ามีความเหมาะสมและสอดคล้องกับข้อมูลในการศึกษา 2 มิติ

1.2.3 ในขั้นตอนของการกำหนดค่าโอกาสการเดาของข้อสอบ (C) ผู้วิจัยพบว่าเมื่อนำผลการตอบข้อสอบที่ได้จากการคำนวณค่าพารามิเตอร์ที่แท้จริง ไปประมาณค่าพารามิเตอร์โดยใช้โปรแกรม BILOG พบว่ามีแนวโน้มที่จะประมาณค่าสูงกว่าความเป็นจริง (Overestimate) ซึ่งส่งผลกระทบต่อความเที่ยงและความตรงของข้อมูล ดังนั้นในการกำหนดค่าผู้วิจัยมีข้อเสนอแนะในการเลือกใช้ 3 แนวทาง โดยแนวทางแรก กำหนดค่า C จากการแจกแจงแบบ $\beta(a,b)$ ซึ่งในการกำหนดค่า a และ b ขึ้นอยู่กับการจำลองข้อมูลในแต่ละครั้ง โดยมีเกณฑ์ในการเลือกใช้ คือค่า c ที่ได้ในทุกข้อไม่ควรเกิน 0.2 แนวทางที่สอง กำหนดค่า C เท่ากับ 0.25 เนื่องจากสอดคล้องกับสถานการณ์จริงมากกว่าแนวทางแรก แต่ดังที่ได้กล่าวมาแล้วข้างต้นว่าการนำข้อมูลไปประมาณค่าพารามิเตอร์มักมีค่าสูงกว่าความเป็นจริงและส่งผลต่อค่าความเที่ยงของข้อสอบ ซึ่งถ้าสถานการณ์การจำลองข้อมูลที่มีจำนวนข้อน้อยเกินไปจะไม่เหมาะสมกับการกำหนดค่า C ดังกล่าว ดังนี้วิธีนี้น่าจะเหมาะสมกับกรณีที่ข้อสอบมีจำนวนข้อ

มากพอ แนวทางที่ 3 การนำค่า C ที่ได้จากการเก็บข้อมูลจริงมาเป็นข้อมูลพื้นฐานเนื่องจากสอดคล้องกับสถานการณ์จริงมากกว่าการกำหนดค่า C รูปแบบอื่น แต่วิธีดังกล่าวทำได้ยากในทางปฏิบัติ

1.2.4 ในขั้นตอนการตรวจสอบแบบแผนการตอบข้อสอบแบบสองค่า (0, 1) ควรมีการพิจารณาให้ผู้สอบแต่ละคนมีโอกาสในการตอบข้อสอบได้ถูกไม่ต่ำกว่าร้อยละ 20 ของข้อสอบทั้งหมด สำหรับนำไปใช้กับแบบสอบเลือกตอบ 4 ตัวเลือก และร้อยละ 25 สำหรับนำไปใช้กับแบบสอบเลือกตอบ 5 ตัวเลือก ถ้าไม่เป็นไปตามเงื่อนไขดังกล่าวควรตัดข้อมูลทิ้งและทำการสร้างข้อมูลใหม่

1.2.5 ในขั้นตอนของการประมาณค่าพารามิเตอร์ (Calibration) โดยใช้โปรแกรม TESTFACT โดยการกำหนดจำนวนรอบเพื่อทำการประมาณค่าความยากง่ายและค่าอำนาจจำแนกของข้อสอบเป็นองค์ประกอบที่สำคัญอย่างหนึ่ง จากการศึกษาพบว่าจำนวนรอบไม่ควรต่ำกว่า 100 รอบ ซึ่งถ้ากำหนดจำนวนรอบน้อยเกินไป จะเกิดความคลาดเคลื่อนสูง โดยพิจารณาได้จากค่าเฉลี่ยเศษที่เหลือ (Mean Residual) ที่มีค่ามากกว่า 0.05

1.2.6 การวิจัยครั้งนี้ผู้วิจัยเลือกใช้โปรแกรม MATLAB เกือบทุกขั้นตอนของการศึกษา ตั้งแต่ขั้นตอนของการสร้างข้อมูล (Generate Data) ไปจนถึงขั้นตอนการเชื่อมโยงคะแนนและการตรวจสอบคุณภาพของการเชื่อมโยงคะแนนเนื่องจากสามารถวิเคราะห์ได้ในคราวเดียวกัน อย่างไรก็ตามการใช้โปรแกรมดังกล่าวจำเป็นที่จะต้องอาศัยผู้ที่มีความรู้ความเชี่ยวชาญเฉพาะด้านในการพัฒนาโปรแกรม โดยเฉพาะอย่างยิ่งวิธี M จำเป็นต้องใช้โปรแกรมดังกล่าว แต่สำหรับวิธี NOP สามารถคำนวณโดยใช้โปรแกรม EXCEL ได้โดยตรง แต่จะต้องระมัดระวังในขั้นตอนของการคำนวณค่าระหว่างเวคเตอร์หรือเมตริกซ์ เพราะจะทำให้ผลที่ได้เกิดข้อผิดพลาดได้ง่าย

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

2. ข้อเสนอแนะสำหรับการวิจัยในอนาคต

2.1 เนื่องจากการศึกษาในครั้งนี้เป็นการศึกษาภาคตัดขวางที่เน้นการเชื่อมโยงคะแนน 3 ระดับชั้นที่แตกต่างกัน ซึ่งผู้วิจัยได้กำหนดค่าการกระจายความสามารถของผู้สอบ (Ability Distribution) ทั้ง 3 ระดับ โดยกำหนดให้มีค่าเฉลี่ยในระดับที่ 1 2 และ 3 เท่ากับ 0, 0.5 และ 1 ตามลำดับ เพื่อสะท้อนให้เห็นถึงลำดับขั้นพัฒนาการของผู้เรียน อย่างไรก็ตามการกำหนดโดยใช้ค่าดังกล่าวอาจไม่สอดคล้องกับสถานการณ์ที่ศึกษา แนวทางหนึ่งสำหรับการศึกษาต่อไปคือ การนำค่ากระจายความสามารถของผู้สอบมาจากการเก็บข้อมูลจริง (Real Data) มาทำการวิเคราะห์แล้วจึงทำการประมาณค่าพารามิเตอร์เพื่อนำไปสู่ขั้นตอนการเชื่อมโยงคะแนนต่อไป

2.2 เนื่องจากการศึกษาในครั้งนี้เป็นการศึกษาเพียง 2 มิติเท่านั้น จึงไม่พบความแตกต่างอย่างเด่นชัดระหว่างวิธี M และ วิธี NOP ดังนั้นในการศึกษาต่อไปควรเปรียบเทียบในมิติที่สูงขึ้น รวมถึงการศึกษาในครั้งนี้มีการศึกษาระดับความสัมพันธ์ของมิติความสามารถเพียง 3 ระดับ ซึ่งผลการวิจัยยังขาดระดับความสัมพันธ์อีกหลายระดับ เช่น 0.3 0.5 0.7 และ 0.9 เป็นต้น โดยเฉพาะอย่างยิ่งเงื่อนไขของความสัมพันธ์ในระดับต่ำ ซึ่งควรมีการศึกษาเพิ่มเติม อย่างไรก็ตามในการเลือกเงื่อนที่จะศึกษาควรพิจารณาให้สอดคล้องกับบริบทเพื่อให้เกิดประโยชน์สูงสุดในการนำไปใช้

2.3 จากผลการวิจัยพบว่ายังขาดความชัดเจนในขั้นตอนของสร้างข้อมูล เนื่องจากยังขาดผลงานวิจัยที่มาช่วยสนับสนุนให้เกิดความชัดเจน อาธิ ขนาดของจำนวนข้อสอบที่เพียงพอต่อการเชื่อมโยงคะแนนที่มีจำนวนมิติแตกต่างกัน การกำหนดค่า C เพื่อให้เกิดความคลาดเคลื่อนในการประมาณค่าน้อยที่สุด การกำหนดขนาดของมุมที่เหมาะสมและสอดคล้องกับโครงสร้างมิติความสามารถเพื่อให้ข้อสอบข้อนั้นอยู่ในมิติที่ต้องการ จำนวนมิติเป็นเท่าใดจึงทำให้ทั้ง 2 วิธี มีคุณภาพของการเชื่อมโยงคะแนนแตกต่างกัน โดยเฉพาะผลที่ได้จากการแปลงค่าพารามิเตอร์ความยาก เป็นต้น ซึ่งประเด็นดังกล่าวควรมีการค้นคว้าวิจัยต่อไปในอนาคต

2.4 เนื่องจากเกณฑ์ที่ใช้ในการพิจารณาคุณภาพของการเชื่อมโยงคะแนนมี 3 เกณฑ์ คือ RMSE BIAS และ ค่าความสัมพันธ์ระหว่างคะแนนที่แปลงได้กับคะแนนเกณฑ์ (CORR) ซึ่งการนำเกณฑ์ CORR ไปใช้จะเหมาะสมเฉพาะการนำไปเปรียบเทียบกับเงื่อนไขที่มีความสัมพันธ์ระหว่างมิติความสามารถในระดับสูง (r=0.8) เท่านั้น ส่วนความสัมพันธ์ระดับอื่นพบว่าทั้งสองวิธีมีค่าไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ดังนั้นถ้าต้องการนำเกณฑ์ CORR ไปใช้ในการวิจัยครั้งต่อไป เหมาะสำหรับการศึกษาเพื่อให้ได้สารสนเทศว่าคะแนนที่แปลงได้กับคะแนนที่เกณฑ์สัมพันธ์กันมากหรือน้อย แต่ไม่เหมาะสมสำหรับนำไปใช้ในการเปรียบเทียบ

คุณภาพระหว่างวิธีการเชื่อมโยงคะแนนเพราะให้คุณภาพที่ไม่ต่างกัน ซึ่งควรเลือกใช้เฉพาะกรณีที่อยู่ในเงื่อนไขที่มีความสัมพันธ์ระหว่างมิติความสามารถในระดับสูงเท่านั้น

2.5 การวิจัยครั้งนี้ผู้วิจัยเลือกใช้โปรแกรม MATLAB ในการพัฒนาโปรแกรมเกือบทุกขั้นตอนของการศึกษา โดยเฉพาะในขั้นตอนของการสร้างข้อมูล (Generate Data) และแบบแผนการตอบข้อสอบของนักเรียน เพื่อให้ตรงกับเงื่อนไขที่ศึกษา คือโครงสร้างมิติความสามารถ ถือว่าเป็นการเลือกใช้โปรแกรมที่เหมาะสมเนื่องจากข้อมูลโดยส่วนใหญ่ต้องอาศัยการคำนวณในเชิงเมตริซ์ และต้องมีการกำหนดเงื่อนไขโครงสร้างมิติความสามารถที่แตกต่างกัน 2 ลักษณะ คือ โครงสร้างที่ซับซ้อนและไม่ซับซ้อน อย่างไรก็ตามการใช้โปรแกรมดังกล่าวมีข้อจำกัดคือผู้ใช้ต้องมีความรู้และเชี่ยวชาญโดยเฉพาะ สำหรับการศึกษาครั้งต่อไปถ้าต้องการจำลองข้อมูลที่ไม่ได้ให้ความสำคัญกับโครงสร้างมิติความสามารถแต่สนใจในเงื่อนไขการจำลองข้อมูลอื่นๆ เช่น ความสัมพันธ์ระหว่างมิติความสามารถ การกำหนดค่าพารามิเตอร์ข้อสอบในรูปแบบต่างๆ ทั้งในรูปแบบการกำหนดค่าคงที่ให้กับค่าพารามิเตอร์ รวมถึงกำหนดให้ค่าพารามิเตอร์มีการแจกแจงลักษณะต่างๆ สามารถนำโปรแกรมสำเร็จรูปมาใช้ได้ เช่น โปรแกรม *WinGen* ที่พัฒนาโดย Han & Hambleton ในปี 2007 ซึ่งเป็นโปรแกรมสำเร็จรูปที่สะดวกและง่ายต่อการนำไปใช้

2.6 เนื่องจากงานวิจัยครั้งนี้ ในการทำซ้ำแต่ละครั้งใช้เวลานานและหลายขั้นตอนผู้วิจัยจึงทำการทำซ้ำเพียง 20 ครั้ง ในการศึกษาครั้งต่อไปควรเพิ่มจำนวนในการทำซ้ำ เช่น 100 ครั้ง เพื่อให้ผลการวิจัยมีความถูกต้องและน่าเชื่อถือมากยิ่งขึ้น

2.7 เนื่องจากงานวิจัยนี้มีข้อจำกัดที่ศึกษาเฉพาะจากสถานการณ์การจำลองข้อมูล ในการศึกษาครั้งต่อไปควรมีการเปรียบเทียบกับการศึกษาโดยใช้ข้อมูลจริง เพื่อให้สามารถเลือกใช้ได้สอดคล้องกับสภาพจริงเพิ่มมากขึ้น

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

รายการอ้างอิง

ภาษาไทย

กัลยา วาณิชย์บัญชา. (2550). การวิเคราะห์ข้อมูลหลายตัวแปร. พิมพ์ครั้งที่ 2. กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

คณะกรรมการการศึกษาแห่งชาติ, สำนักงาน. (2545). *พระราชบัญญัติการศึกษาแห่งชาติ พ.ศ. 2542 (ฉบับแก้ไข 2545)*. กรุงเทพมหานคร: พริกหวานกราฟฟิก.

ชยุตม์ ภิรมย์สมบัติ. (2547). คุณสมบัติของตัวประมาณค่าความเข้มของอิทธิพล: การเปรียบเทียบระหว่างทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบ.วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาวิจัยและจิตวิทยาการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

ณัฏฐภรณ์ หลาวทอง และ สกล ซื่อธนาพรกุล. (2550). ความคิดเห็นของนิสิตต่อมาตรฐานวิชาชีพครู ด้านความรู้ ในหลักสูตรครุศาสตรบัญฑิต จุฬาลงกรณ์มหาวิทยาลัย: การวิเคราะห์เปรียบเทียบ และการวิเคราะห์พหุระดับ. (เอกสารอัดสำเนา).

พรพิมล นาคเวช. (2537). *การศึกษาคุณภาพการเทียบมาตราในแนวตั้งโดยใช้ทฤษฎีตอบสนองข้อสอบ*. วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

พัชรี จันทร์เพ็ง. (2547). *การศึกษาพัฒนาการความสามารถทางคณิตศาสตร์ของนักเรียนโดยวิธีการปรับเทียบแนวตั้งตามวิธีทฤษฎีตอบสนองข้อสอบร่วมกับวิธีการปรับเทียบเชิงเส้นตรง*. วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาวิจัยและจิตวิทยาการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

พิชัย ละแมนชัย. (2539). *ขนาดกลุ่มตัวอย่างขั้นต่ำสำหรับการปรับเทียบคะนนในแนวระดับตามทฤษฎีการตอบสนองข้อสอบที่มีแบบแผนการปรับเทียบและความยาวแบบสอบแตกต่างกัน*. วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

ภัทราพร เกษสังข์. (2546). *การศึกษาพัฒนาการความสามารถทางคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาตอนต้นโดยการปรับเทียบคะแนนในแนวตั้งที่ใช้วิธีการที่เหมาะสม*. วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒประสานมิตร.

วรนุช แหยมแสง. (2536). *การพัฒนากระบวนการตรวจสอบความเป็นเอกมิติของแบบสอบ*. วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

วรรณ์ดี แสงประทีปทอง. (2538). *การศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนวิชาภาษาไทยของนักเรียนชั้นมัธยมศึกษาตอนต้น โดยการปรับเทียบคะแนนตามแนวตั้งตามวิธีทฤษฎีตอบสนองข้อสอบแบบจำลองโลจิสติกสามพารามิเตอร์*. วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒประสานมิตร.

ศิริชัย กาญจนวาสี. (2541). *การปรับเทียบคะแนนระหว่างแบบสอบ.* กรุงเทพมหานคร: คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

ศิริชัย กาญจนวาสี. (2550). *ทฤษฎีการทดสอบแนวใหม่ (Modern Test Theories).* พิมพ์ครั้งที่ 3. กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

ศิริชัย กาญจนวาสี. (2550). *ทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory).* พิมพ์ครั้งที่ 5. กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

ศึกษาธิการ, กระทรวง, สำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ, สำนักนิเทศและพัฒนามาตรฐานการศึกษาแห่งชาติ. (2544). *พัฒนาการของคุณภาพนักเรียนประถมศึกษาและแนวทางการประเมิน.* กรุงเทพมหานคร: โรงพิมพ์การศาสนา.

สุนิสา จุ้ยม่วงศรี. (2546). *การศึกษาผลของการปรับเทียบคะแนนข้อสอบที่ตรวจให้คะแนนแบบหลายค่า.* วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร.

สุมาลี ชูศรี. (2536). การพัฒนารูปแบบปฏิสัมพันธ์ของการสอนอ่านเข้าใจความ. วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาจิตวิทยาการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

สุวิมิล ติรการนันท์. (2548). *การตรวจสอบความเป็นเอกมิติ (Examining The Unidimensionality) (ปรับปรุงใหม่).* กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย

เอมอร จังศิริพรปกรณ์. (2548). *การเปรียบเทียบคุณภาพของแบบสอบระหว่างแบบสอบเลือกตอบที่มีจำนวนตัวถูกตัวเดียวกับตัวถูกหลายตัวเมื่อตรวจด้วยวิธีการให้คะแนนความรู้บางส่วน.* กรุงเทพมหานคร: โรงพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.

<u>ภาษาอังกฤษ</u>

Ackerman, T. A. (1991). The Use of Unidimensional Parameter Estimates of Multidimensional Item in Adaptive Testing. *Applied Psychological Measurement*, 15, 12-24

Ackerman, T. A. (1992). A Didactic Explanation of Item Bias, Item Impact, and Item Validity from a Multidimensional Perspective. *Journal of Educational Measurement*, 29, 67-91.

Ackerman, T. A. (1994). Using Multidimensional Item Response Theory to Understand What Items and Tests are Measuring. *Applied Psychological Measurement*, 18, 255-278.

Ackerman, T. A. (1996). Graphical Representation of Multidimensional Item Response Theory Analyses. *Applied Psychological Measurement*, 20, 311-329.

Ackerman, T. A., Gierl, M. J., & Walker, C. M. (2003). Using Multidimensional Item Response Theory to Evaluate Educational and Psychological Tests. *Educational Measurement: Issues and Practice*, 22, 37-53.

Ackerman, T. A., Gierl, M. J., & Walker, C. M. (2003). Using Multidimensional Item Response Theory to Evaluate Educational and Psychological Tests. *Educational Measurement: Issues and Practice*. Fall

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

Angoff, W. H. (1971). Scales, Norm and Equivalent Score. In R.L. Throndike(Ed.), *Education Measurement*. (2nd ed.). Washington, D.C.: American Council on Education, 508-600.

Ayerve, R. I. (1992). The Effectiveness of The Equipercentile Method and IRT Three Parameter Model on Vertical Equating under Verying Condition of Sample Size, Test Length and Anchor Test Length: A Simulation Study. Doctoral Dissertation. Columbia university. *Dissertation Abstracts International*, 53, 1841

Baker, F. B. (1996). An Investigation of The Sampling Distribution of Equating Coefficients. *Applied Psychological Measurement*, 20, 45-57.

Bereiter, C. (1963). Some Persisting Dilemmas in The Measurement of Change. In C. W. Harris (Ed.), *Problems in measuring change*. Milwaukee, WI: University of Wisconsin Press.

Bergman, L., Eklund, G., Magnusson, D. (1991). *Problems and Methods in Longitudinal Research: Stability and Change*. New York: Cambridge University Press.

Bock, R. D., & Aitkin, M. (1981). Marginal Maximum Likelihood Estimation of Item Parameters: Application of an EM Algorithm. *Psychometrika,* 46, 443-459.

Bock, R. D., Gibbons, R., & Muraki, E. (1998). Full Information Item Factor Analysis. *Applied Psychological Measurement*, 12, 261-270.

Bolt, D. M. (1999). Evaluating The Effects of Multidimensionality on IRT True-Score Equating. *Applied Measurement in Education*, 12, 383-407.

aun, H. L., & Holland, P. W. (1982). Observed-Score Test Equating: A Mathematical Analysis of Some ETS Equating Procedures, In Holland, P.W. and Rubin, D.B (Eds.) *Test Equating*. New York: Academic Press.

Briggs, D. C., & Wilson. M. (2003). An Introduction to Multidimensional Measurement using Rasch Models. *Journal of Applied Measurement*, 4(1), 87-100.

Budescu, D. (1985). Efficiency of Linear Equating as a Function of The Length of The Anchor Test. Journal of Educational Measurement, 22, 13-20.

Camilli, G. Wang, M., & Fesq, J. (1995). The Effect of Dimensionality on Equating The Law School Admission Test. *Journal of Educational Measurement*, 32, 79-96.

Carlson, J. E. (1987). *Multidimensional Item Response Theory Estimation: A Computer Program*. Iowa City IA: American College Testing Program.

Christoffersson, A. (1975). Factor Analysis of Dichotomized Variables. *Psychometrika*, 40, 5–32.

Clauser, B. E., & Mazor, K. M. (1998). Using Statistical Procedures to Identify Differentially Functioning Test Items. *Educational Measurement: Issues and Practice,* 17, 31–44.

Cook, L. L., & Eignor, D. R. (1991). An NCME Instructional Module on IRT Equating Methods. *Educational Measurement: Issues and Practice*, 10, 37-45.

Cook, L., & Eignor, D. R. (1989). Using Item Response Theory in Test Score Equating. *International Journal of Education Research*, 13, 161-173

CTB/McGraw-Hill. (2002). *Technical Bulletin 1 of California Achievement Tests Forms C and D.* Monterey, CA: Author.

Davey, T. C., & Oshima, T. C. (1994). *Linking Multidimensional Item Calibrations*. Paper Presented at The Annual Meeting of The National Council on Measurement in Education, New Orleans.

Davey, T., Nering, M. L., & Thompson, T. (1997). *Realistic Simulation of Item Response Data*. Act Research Report Series ONR 97-4. Iowa City, IA: ACT, Inc.

De Champlain, A. F., & Tang, K. L. (1997) CHIDIM: A FORTRAN Program for Assessing The Dimensionality of Binary Item Responses Based on McDonald's Nonlinear Factor Analytic Model. *Educational and Psychological Measurement*, 57, 174–178.

Dorans, N. J. (2000). Scaling and Equating. In H. Wainer (Ed.), *Computerized Adaptive Testing: A Primer (2nd)*. New Jersey: Lawrence.

Drasgow, F., Levine, M. V., & McLaughlin, M. E. (1991). Appropriateness Measurement for some Multidimensional Test Batteries. *Applied Psychological Measurement*, 15, 171-191.

Embretson, S. E. (1984). A General Latent Trait Model for Response Processes. *Psychometrika*, 40, 175-186.

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

Feuer, M. J., Holland, P. W., Green, B. F., Bertenthal, M. W., & Hemphill, F. C. (1999). *Uncommon Measures: Equivalence and Linkage among Educational Tests*. Washington, DC: National Academy Press.

Flanagan, T. C. (1951). Units, Score and Norms. In EF . Linguist(Ed.) *Education Measurement*. Washing, D.C.: American Council on Education.

Fraser, C., & McDonald, R. (1986). *NOHARM II: A FORTRAN Program for Fitting Unidimensional and Multidimensional Normal Ogive Models of Latent Trait Theory*. Amidale, Australia: University of New England, Center for Behavioral Studies.

Froelich, A. G., & Habing, B. (2001). *Refinements of The DIMTEST Methodology for Testing Unidimensionality and Local Independence*. Paper Presented at The Aannual Meeting of The National Council on Measurement in Education, Seattle, WA.

Glas, C. A. W. (1992). A Rasch Model with A Multivariate Distribution of Ability. In M. Wilson (Ed.), *Objective Measurement: Theory into Practice.* Norwood New Jersy: Ablex, 1, 236-258.

Glass, G. V., & Hopkins, K. D. (1996). *Statistical Methods in Education and Psychology.* (3th ed.). New Jersey: Prentice-Hall Inc.

Glowacki, M. L. (1991). The Analysis of Test Equating Models for The Alabama High School Graduation Examination. Doctoral Dissertation, University of Alabama. *Dissertation Abstracts International*, 52, 1722

Green, P. E. (1976). *Mathematical Tools for Applied Multivariate Analysis*. New York: Academic Press.

Guliksen, H. (1950). *Theory of MentalTtest*. New York: Wiley.

Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. Boston: Kluwer.

Hambleton, R. K., & Rovinelli, R. J. (1986). Assessing The Dimensionality of a Set of Test Items. *Applied Psychological Measurement*, 10, 287–302.

ambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory: Principle and Applications.* (2nd ed). Boston: Kluwer Nijhoff Publishing.

Harris, D. (1989) Comparison of 1, 2, and 3 Parameter IRT Models. *Educational Measurement: Issues and Practice*, 8, 35–41.

Harris, D. J., & Hoover, H. D. (1987). An Application of The Three-parameter IRT Model to Vertical Equating. *Applied Psychological Measurement*, 11, 151-159.

Harwell, M. R., Stone, C. A., Hsu, T., & Kirisci, L. (1996). Monte Carlo Studies in Item Response Theory. *Applied Psychological Measurement*, 20, 101-125.

Hirsch, T. M. (1989). Multidimensional Equating. *Journal of Educational Measurement*, 26, 337-349.

Jodoin, M. G., et al. (2003). Acomparison of Linear, Fixed Common, and Concurrent Paramiter Estimation Equating Procedures in Capturimg Academic Growth. *The Journal of Experimental Education*, 71(3), 229-50

Johnson, E. G., & Carlson, J. (1994). *The NAEP 1992 Technical Report* (Report No. 23-TR-20). Washington, DC: National Center for Education Statistics.

Johnson, E. G., & Owen, E. (1998). Linking The Nation Assessment of Education Progress (NAEP) and The Third International Mathematic and Science Study (Timss): A Technical Report. (Online). U.S.A.: National Center For Education Statistics, *Research and Development Report*. Retrieved August 8, 2006, from: http://www.ETS.ORG.

Kelderman, H. (1994). Objective Measurement with Multidimensional Polytomous Latent Trait Model. *Objective Measurement: Theory into Practice*. Norwood New Jersy: Ablex. 2, 235-243.

Kelderman, H., & Rijkes, C. (1994). Loglinear Multidimensional IRT Models for Polytomously Scored Items. *Psychometrika,* 59, 149–176.

Kim, H. (1994). *New Techniques for The Dimensionality Assessment of Standardized Test Data*. Unpublished Doctoral Dissertation. University of Illinois at Urbana-Champaign.

Knol, D. L. & Berger, M. P. F. (1991). Empirical Comparison Between Factor Analysis and Multidimensional Item Response Models. *Multivariate Behavioral Research*, *26*, 457-477.

Knol, D. L., & Berger, M. P. F. (1991). Empirical Comparison Between Factor Analysis and Multidimensional Item Response Models. *Multivariate Behavioral Research*, 26, 457

Kolen M. L., & Brennan R. L. (2004). *Test Equating, Scaling, and Linking : Methods and Practices*. New York: Springer Science Business Media.

Kolen, M. J. (2001). Linking Assessments Effectively: Purpose and Design. *Education Measurement: Issues and Practice*, 20, 5-9.

Kolen, M. L., & Brennan, R. L. (1995). *Test Equating Method and Practices*. New York: Spring

Leung, S.O. (2003). A Practical Use of Vertical Equating by Combining IRT Equating and Linearing Equating. *Practical Assessment, Research & Evaluation.* Retrieved June 8, 2004, from http://PARE online.net/getven.asp? v=8&n=23

Li, Y. H. (1996). *MDEQUATE (Computer Software)*. Upper Marlboro MD: Author.

Li, Y. H. (1997). An Evaluation of Multidimensional IRT Equating Parameters onto a Target Test Metric (Doctoral Dissertation, University of Maryland, 1997). *Dissertation Abstracts International*, UMI Number 9816494.

Li, Y. H., & Lissitz, R. W. (2000). An Evaluation of The Accuracy of Multidimensional IRT Linking. *Applied Psychological Measurement*. 24(2), 115-138.

Li, Y. H., & Lissitz, R. W. (1998). An Evaluation of Multidimensional IRT Equating Methods by Assessing The Accuracy of Transforming Parameters onto a Taget Test Metric. *Paper presented at The Annual Meeting of The National Council on Measurement in Education*, San Diego, CA.

Lissitz, R. W., & Huynh, H. (2003). Vertical Equating for State Assessment: Issues and Solution in Determination of Adequating Yearly Progress and School Accountability. *Practical Assessment, Research & Evaluation*. Retrieved June 8, (2007), from: http://PARE online.net/getven.asp? v=8&n=10

Lord, F. M. (1980). *Application of Item Response Theory to Practical Testing Problem.* New Jersey: Hillssdale Erlbaun.

Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. New Jersey: Lawrence.

Lord, F. M., & Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Reading: Addison-Welsley Publishing Company.

Ma, X., & Wilkins, J. L. M. (2002). The Development of Science Achievement in Middle and High School: Individual Differences and School Effects. *Evaluation Review,* 26, 395-417.

Marion, S. (2004). *Validity Considerations when Measuring Growth*. Center for Assessment. The 6th Annual Reidy Interaction Lecture Series.

Mark, J. G. (2005). Using Dimensionality-Based DIF Analyses to Identify and Interpret Constructs That Elicit Group Differences. *Educational Measurement: Issues and Practice*, 24(1), 3-14

Martineau, J. A., & Reckase, M. D. (2006). *Practical Multidimensional Vertical Equating Based on Accurate Assessment of Dimensionality.* Paper Presented at The 36th Annual National Conference on Large-Scale Assessment of The Council of Chief State School Officers (CCSSO), San Francisco, CA.

McDonald, R. P. (1999). *Test Theory: A Unified Treatment.* New Jersey: Lawrence Erlbaum Associates.

McDonald, R. P. (1997). Normal-Ogive Multi-dimensional Model. In V. J. van der Linden & R. K. Hambleton, (Eds.), *Handbook of Modern Item Response Theory*. New York: Springer, 257-269.

McDonald, R. P. (2000). A Basis for Multidimensional Item Response Theory. *Applied Psychological Measurement,* 24, 99–114.

McKinley, R. L., & Mils, C. N. (1985). A Compation of Several Goodness-of-Fit Statistics. *Applied Psychological Measurement*, 9, 49-57.

McKinley, R. L., & Reckase, M. D. (1983). *An Extension of The Two-Parameter Logistic Model to The Multidimensional Latent Space (ACT Research Report ONR83-2*). Iowa City, IA: American College Testing.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement*. (3 rd ed.). New York: American Council on Education/Macmillan.

Messick, S. (1995). Validation of Psychological Assessment: Validation of Inferences form Person's Responses and Performances as Scientific Inquiry into Score Meanings. *American Psychologist, 50*, 741-749.

Miller, T. R., & Hirsch, T. M. (1992). Cluster Analysis of Angular Data in Application of Multidimensional Item Response Theory. *Applied Measurement in Education*, 5, 193-211.

Min, K. S. (2003). *The Impact of Scale Dilation on The Quality of The Linking of Multidimensional Item Response Theory Calibrations*. Unpublished Dissertation, Michigan State University, East Lansing, MI.

Mislevy, R. J. (1986). Recent Developments in The Factor Analysis of Categorical Variables. *Journal of Educational Statistics*, 11, 3–31.

Mulaik, S. A. (1972). *The Foundations of Factor Analysis*. New York: McGraw-Hill.

Muraki, E., & Engelhard, G. (1985). Full-Information Item Factor Analysis: Applications of EAP Scores. *Applied Psychological Measurement*, 9, 417-430.

Nandakumar, R. (1991). Traditional Dimensionality Versus Essential Dimensionality. *Journal of Educational Measurement*, 28, 99–117.

Nandakumar, R., & Stout, W. (1993). Refinement of Stout's Procedure for Assessing Latent Trait Unidimensionality. *Journal of Educational Statistics*, 18, 41–68.

Nandakumar, R., Yu, F., Li, H., & Stout, W. (1998). Assessing Unidimensionality of Polytomous Data. *Applied Psychological Measurement*, 22, 99–115.

O'Brien, M. l., & Tohn, D. (1984). *Applying and Evaluating Rasch vertical Equating Proceduce for Out-of-Level Testing.* Paper Presented at The Annual Meeting of The Eastern Educational Research Association , West Palm Beach, FL.

Oshima, T.C., Davey, T., & Lee, S. (2000). Multidimensional Linking: Four Practical Approaches. *Journal of Educational Measurement*, 37(4), 357-373.

Petersen et al., (1982). A Test of The Adequacy of Linear Score Equating Methods, In R.L. Linn (ED.) *Educational Measurement*. (3rd ed.) New York: Macmillan

Petersen, N. S., Kolen, M. J., & Hoover, H. D. (1989). Scaling, Norming and Equating. In R. L. Linn (Ed.) *Educational Measurement* . New York: Macmillan.

Rackase, M. D. (1997). The Past and Future of Multidimensional Item Response Theory. *Applied Psychological Measurement*, 21(1), 25-36

Reckase, M. D. (1995). A Linear Logistic Multidimensional Model for Dichotomous Item Response Data. In W. J. van der Linden and Hambleton (Ed.), *Handbook of Modern Item Response Theory*. New York: Springer.

Reckase, M. D. (1997). A Linear Logistic Multidimensional Model for Dichotomous Item Response Data. In W. J. van der Linden and R. K. Hambleton (Eds.) *Handbook of Modern Item Response Theory*. New York: Springer.

Reckase, M. D. (1985). The Difficulty of Test Items That Measure More Than One Ability. *Applied Psychological Measurement*, 9, 401–412.

Reckase, M. D., & Martineau, J. A. (2004). *The Vertical Scaling of Science Achievement Tests*. Paper Commissioned by The Committee on Test Design for K-12 Science Achievement Center for Education National Research Council.

Reckase, M. D., & McKinley, R. L. (1991). The Discriminating Power of Items That Measure More Than One Dimension. *Applied Psychological Measurement*, 15, 361-373.

Reckase, M. D., Ackerman, T. A., & Carlson, J. E. (1988). Building a Unidimensional Test Using Multidimensional Items. *Journal of Educational Measurement*, 25(3), 193-203.

Reskase, M. D. (1997). A Linear Logistic Multisimensional Model for Dichotomous Item Response Data. In W.J. Linder and R.K. Hambleton (Eds.), *Handbook of Modern Item Response Theory*. New York: Springer.

Reskase, M. D. (2007). Multidimensional Item Response Theory. (in press).

Reskase, M. D., Thompson, T. D., & Nering, M. (1997). *Identifying Similar Item Content Clusters on Multiple Test Forms.* Paper Presented at The Annual Meeting of The Psychometric Society, Gatliburg TN.

Reynolds, C. R., & Wilson, V. L. (1984). Standardized Grade Equivalents: Really! No. Well, Sort of, but They are More Confusing than Hlpful. *Journal of learning Disability*, 17(6), 326-327.

Rogers, H. J. (1999). Guessing in Multiple Choice Tests. In G. N. Masters & J. P. Keeves (Eds.), *Advances in measurement in educational research and assessment*. Amsterdam: Pergamom.

Rotou, O., Headrick, T. C. & Elmore, P. B. (2002). A Proposed Number Correct Scoring Procedure based on Classical True-Score Theory and Multidimensional Item Response Theory. *International Journal of Testing, 2*, 131-141.

Roussios, L. A., Stout, W. F., & Marden, J. I. (1998). Using New Proximity Measures with Hierarchical Cluster Analysis to Detect Multidimensionality. *Journal of Education Measurement*, 35, 1-30.

Schulz, E. M., & Nicewander, W. A. (1997). Grade Equivalent and IRT Representations of Growth. *Journal of Educational Measurement*, 34 (4), 315-331.

Skaggs, G., & Lissitz, R. W. (1988). Effect of Examinee Ability on Test Equating Invariance. *Applied Psychological Measurement*, 12(1), 69-82.

Slinde, J. A., & Linn, R. L. (1978). Vertical Equating Test: Fact or Phantom. *Journal of Educational Measurement,* 14, 23-31.

Spray, J. A. et al. (1990). *Comparison of Two Logistic Multidimensional Item Response Theory Models (Research Report ONR 90-8)*. Iowa City, IA: American College Testing.

Stevens, S. S. (1951). *Handbook of experimental psychology*. New York: Wiley.

Stout, W. (1987). A Nonparametric Approach for Assessing Latent Ttrait Unidimensionality. *Psychometrika*, 52, 589–617.

Sympson, J. B. (1978). A Model for Testing with Multidimensional Items. In D. J. Weiss (Ed.), *Proceedings of The 1977 computerized adaptive testing conference*, 82–103.

Tam. H. P. & Li, Y. H. (1997). *Is The Use of The Difference Likelihood Ratio Chi-Squre Statistic for Comparing Nesred IRT Models Justifiable?.* Paper Presented at The Manual Meeting of American Education Research Association. Chicago, IL.

Thompson, T. D., Nering, M. & Davey, T. (1997). *Multidimensional IRT Scale Linking without Common Items or Common Examinees.* Paper Presented at The Annual Meeting of The Psychometric Society, Gatliburg TN.

Thorndike R. L., & Hagen E. P. (1977). *Measurement and Evaluation in Psychology and Education* (4th ed.). New York: Wiley.

Thum, Y. M. (2003). *No Child Left Behind: Methodological Challenges & Recommendations for Measuring Adequate Yearly Progress*. Los Angeles, CA: Center for The Study of Evaluation, University of California, Los Angeles.

Torre, J., & Patz, R. J. (2006). Making the What We Have: A Practical Application of Multidimensional Item Response Theory in Test Scoring. *Journal of Education and Behavioral Statistics*, 30(3): 295-311

Traub, R. E. (1983). A Priori Considerations in Choosing an Item Response Model. In R. K. Hambleton (Ed.), *Applications of Item Response Theory*. Vancouver, BC: Educational Research Institute of British Columbia.

Vale, D. (1986). Linkink Item Parameters onto a Common Scale. *Applied Phychological Measurement*, 10(4), 333-344.

Van Der Linden, W. J., & Hambleton, R. K. (1997). Handbook of Modern Item Response Theory. New York: Springer.

Walker, C. M., & Beretvas, S. N. (2000). *Using Multidimensional Versus Unidimensional Ability Estimates to Determine Student Proficiency in Mathematics*. Paper Presented at The Meeting of The American Educational Research Association, New Orleans, LA.

Whitely, S. E. (1980). Multicomponent Latent Trait Models for Ability Tests. *Psychometrica*, 45, 479-494.

Wiley, A. (1999). An Investigation into Two Model For Equating Examinations with Multiple Item Formats (Anchor Testing), Doctoral dissertation, Fordham University. *Dissertation Abstracts International,* 59(7-A).

Wilson, D., Wood, R., & Gibbons, R. D. (1991). *TESTFACT: Tests Scoring, Item Statistics, and Item Factor Analysis*. Mooresville, ID: Scientific Software

Wingersky, M. S., & Lord, F. M. (1984). An Investigation of Methods for Reducing Sampling Error in Curtain IRT Procures. *Applied Phychological Measurement*, 8, 347-364.

Wright, B. D. (1968). *Sample-free Test Calibration and Person Measurement*. Proceedings of The 1967 Invitational Conference on Testing Problems. Princeton, New Jersey: Educational Testing Service.

Yang, W. (1997). The effects of Common-Item Test Equating, Doctoral dissertation, Michigan State University. *Dissertation Abstracts International*, 59(7-A).

Yeh, C. C. (2007). The Effect of Guessing on Assessing Dimensionality in Multiple-Choice Test: A Monte Carlo Study with Application. *Doctoral Dissertation*. University of Pittsburgh.

Yen, W. M., & Burket, G. R. (1997). Comparison of Item Response Theory and Thurstone Methods of Vertical Scaling. *Journal of Educational Measurement*, 34(4), 293-313.

Yon, H. (2006). Multidimensional Item Response Theory Approaches to Vertical Scaling. *Doctoral Dissertation*. Michigan State University

Yon, H., Sullivan, C., & Simpson, M. A. (2007). *The Effects of Non-orthogonal Rotation in Multidimensional Linking of Tests Displaying Construct Shift.* Paper Presented at The Annual Meeting of The American Educational Research Association (AERA) and The National Council on Measurement in Education (NCME) held Between April 9, 2007 - April 13, 2007 in Chicago.

Zhang, J. (1997). *The Composite Best Measure by a Multidimensional Test Score*. Paper Presented at The Annual Meeting of The Psychometric Society, Gatliburg TN.

Zhang, J., & Stout, W. (1999). The Theoretical Detect Index of Dimensionality and Its Application to Approximate Simple Structure. *Psychometrika, 64,* 231-249.

ภาคผนวก

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ก

รายชื่อผู้เชี่ยวชาญในการตรวจสอบโปรแกรมที่ผู้วิจัยพัฒนาขึ้น

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

รายชื่อผู้เชี่ยวชาญ

1. ว่าที่ร้อยโทภณัฐ ก้วยเจริญพาณิชก์	อาจารย์ประจำภาควิชาคณิตศาสตร์ สถิติ และคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี
2. อาจารย์ สังวรณ์ งัดกระโทก	อาจารย์ประจำศูนย์วิชาการประเมินผล มหาวิทยาลัยสุโขทัยธรรมาธิราช

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ข

ตัวอย่างคำสั่ง MATLAB สำหรับจำลองแบบแผนการตอบของผู้สอบ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตัวอย่างคำสั่ง MATLAB สำหรับจำลองแบบแผนการตอบ
ตามโครงสร้างที่ไม่ซับซ้อน (APSS) และโครงสร้างที่ซับซ้อน (MS)

```
% INPUT PROGRAM

% Program  binorm.m

% This MATLAB Code simulates a bivariate normal distribution


% with read in means and standard deviations and correlation coefficient

% กำหนด ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานในแต่ละมิติ
% xm=input('mean of the x variate:    ')
xm = 0.0;

%ym=input('mean of the y variate:    ')
ym = 0.0;

%xsig=input('standard deviation of the x variate:    ')
xsig = 1.0;

%ysig=input('standard deviation of the y variate:    ')
ysig = 1.0;

% กำนหนดระดับความสัมพันธ์ของมิติความสามารถระหว่าง θ1 และ θ2
%s=1;
%while s>0;
%rho=input('correlation coefficient between -1,1:    ')
rho = 0.00;

theta1=[ ];
theta2=[ ];

%  กำนหนดขนาดข้อสอบร่วมและขนาดผู้สอบที่ต้องการ
%  ในที่นี้ใช้ข้อสอบ 20 ข้อ กลุ่มตัวอย่าง 2,000 คน
NM = 2000;
NI = 30;
```

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

```
% จำลองข้อมูลแบบ bivariate normal distribution ตามเงื่อนไขดังกล่าว
i=0;
while (i < NM),
a=randn;
x1=xm+xsig*a;
y1=ym+rho*ysig*a+ysig*randn*sqrt(1-rho*rho);
if (x1>-3)&(x1<3)
  if (y1>-3)&(y1<3)
        theta1=[theta1 x1];
        theta2=[theta2 y1];
        i=i+1;
   end;
end;
end;

% แสดงแฟ้มข้อมูลค่าความสามารถที่แท้จริงของผู้สอบในมิติที่ 1 และ 2
disp('theta1 :')
theta1(:)
disp('theta2 :')
theta2(:)

% อ่านแฟ้มข้อมูลค่าความสามารถที่แท้จริงจาก INPUT FILE (ใช้ในกรณีที่นำเข้าจากแฟ้มข้อมูลอื่น)
    %fname = ['theta_2.txt'];
    %fid = fopen(fname);
    %NM = fscanf(fid,'%d\n',1);
    %a = fscanf(fid,'%g %g\n',[2 inf]);
    %fclose(fid);
    %theta1 = a(1,:);
    %theta2 = a(2,:);
```

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

% การสุ่มมุมแบบ uniform distribution ตามโครงสร้าง APSS เพื่อคำนวณค่าอำนาจจำแนกของข้อสอบ

```
angle = [ ];

h = 30.0/(NI-2)
angle(1,1) = 0;
angle(NI/2  ,1) = 15;
angle(NI/2 +1 ,1) = 75;
angle(NI ,1) = 90;
for i=2:NI/2 -1
   angle(i,1) = angle(i-1,1) + h;
end
for i=NI/2 +2:NI-1
   angle(i,1) = angle(i-1,1) + h;
end
for i=1:NI
   angle(i,2) = 90.0 - angle(i,1);
end

angle(1,1) = angle(1,1) + 1;
angle(1,2) = 90.0 - angle(1,1);
angle(NI,1) = angle(NI,1) - 1;
angle(NI,2) = 90.0 - angle(NI,1);
```

% การสุ่มมุมแบบ uniform distribution ตามโครงสร้าง MS เพื่อคำนวณค่าอำนาจจำแนกของข้อสอบ

```
%angle = [ ];

%h1 = 15.0/6.0;
%angle(1,1) = 0.0;
%angle(7,1) = 15.0;

%h2 = 15.0/8.0;
%angle(8,1) = 25.0;
%angle(15,1) = 40.0;

%h3 = 15.0/8.0;
%angle(16,1) = 50.0;
%angle(23,1) = 65.0;

%h4 = 15.0/6.0;
%angle(24,1) = 75.0;
```

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

```
%angle(NI ,1) = 90.0;

%for i=2:6
%    angle(i,1) = angle(i-1,1) + h1;
%end
%for i=9:14
%    angle(i,1) = angle(i-1,1) + h2;
%end
%for i=17:22
%    angle(i,1) = angle(i-1,1) + h3;
%end
%for i=25:NI-1
%    angle(i,1) = angle(i-1,1) + h4;
%end
%for i=1:NI
%   angle(i,2) = 90.0 - angle(i,1);
%end

%angle(1,1) = angle(1,1) + 0.5;
%angle(1,2) = 90.0 - angle(1,1);

%angle(NI,1) = angle(NI,1) - 0.5;
%angle(NI,2) = 90.0 - angle(NI,1);
```

% อ่านแฟ้มข้อมูลขนาดของมุมในแต่ละข้อจาก INPUT FILE (ใช้ในกรณีที่นำเข้าจากแฟ้มข้อมูลอื่น)

```
%     fname = ['angle.txt'];
%     fid = fopen(fname);
%     NI = fscanf(fid,'%d\n',1);
%     a = fscanf(fid,'%g %g\n',[2 inf]);
%     fclose(fid);
%     angle(:,1) = a(1,:)';
%     angle(:,2) = a(2,:)';
```

%หมายเหตุ: %การสุ่มมุมแบบ uniform distribution ตามโครงสร้าง APSS และ MS ให้เลือก
%เฉพาะโครงสร้างโครงสร้างหนึ่งเท่านั้นในการจำลองข้อมูลแต่ละครั้ง
%ถ้าไม่ต้องการให้โปรแกรมประมวลผลในโครงสร้างใด ให้ใส่เครื่องหมาย %
%ไว้ข้างหน้าตัวอักษรทุกบรรทัดของโครงสร้างนั้น

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

% การกำหนดค่าอำนาจจำแนกรวมและค่าความยากง่ายรวมในแต่ละข้อของข้อสอบร่วม 30 ข้อ
% ของนักเรียนระดับที่ 1 ทำแบบสอบระดับที่ 2

MDISC(1)=1.3;
MDISC(2)=1.2;
MDISC(3)=1.0;
MDISC(4)=0.8;
MDISC(5)=0.6;

MDISC(6)=1.3;
MDISC(7)=1.2;
MDISC(8)=1.0;
MDISC(9)=0.8;
MDISC(10)=0.6;

MDISC(11)=1.3;
MDISC(12)=1.2;
MDISC(13)=1.0;
MDISC(14)=0.8;
MDISC(15)=0.6;

MDISC(16)=1.3;
MDISC(17)=1.2;
MDISC(18)=1.0;
MDISC(19)=0.8;
MDISC(20)=0.6;

MDISC(21)=1.3;
MDISC(22)=1.2;
MDISC(23)=1.0;
MDISC(24)=0.8;
MDISC(25)=0.6;

MDISC(26)=1.3;
MDISC(27)=1.2;
MDISC(28)=1.0;
MDISC(29)=0.8;
MDISC(30)=0.6;

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

```
MDIFF(1)=3.0;
MDIFF(2)=0.5;
MDIFF(3)=1.5;
MDIFF(4)=2.5;
MDIFF(5)=0.0;

MDIFF(6)=3.0;
MDIFF(7)=0.5;
MDIFF(8)=1.5;
MDIFF(9)=2.5;
MDIFF(10)=0.0;

MDIFF(11)=3.0;
MDIFF(12)=0.5;
MDIFF(13)=1.5;
MDIFF(14)=2.5;
MDIFF(15)=0.0;

MDIFF(16)=3.0;
MDIFF(17)=0.5;
MDIFF(18)=1.5;
MDIFF(19)=2.5;
MDIFF(20)=0.0;

MDIFF(21)=3.0;
MDIFF(22)=0.5;
MDIFF(23)=1.5;
MDIFF(24)=2.5;
MDIFF(25)=0.0;

MDIFF(26)=3.0;
MDIFF(27)=0.5;
MDIFF(28)=1.5;
MDIFF(29)=2.5;
MDIFF(30)=0.0;
```

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

% การกำหนดค่าอำนาจจำแนกรวมและค่าความยากง่ายรวมในแต่ละข้อของข้อสอบร่วม 30 ข้อ
% ของนักเรียนระดับที่ 2 ทำแบบสอบระดับที่ 2

MDISC(1)=1.3;
MDISC(2)=1.2;
MDISC(3)=1.0;
MDISC(4)=0.8;
MDISC(5)=0.6;

MDISC(6)=1.3;
MDISC(7)=1.2;
MDISC(8)=1.0;
MDISC(9)=0.8;
MDISC(10)=0.6;

MDISC(11)=1.3;
MDISC(12)=1.2;
MDISC(13)=1.0;
MDISC(14)=0.8;
MDISC(15)=0.6;

MDISC(16)=1.3;
MDISC(17)=1.2;
MDISC(18)=1.0;
MDISC(19)=0.8;
MDISC(20)=0.6;

MDISC(21)=1.3;
MDISC(22)=1.2;
MDISC(23)=1.0;
MDISC(24)=0.8;
MDISC(25)=0.6;

MDISC(26)=1.3;
MDISC(27)=1.2;
MDISC(28)=1.0;
MDISC(29)=0.8;
MDISC(30)=0.6;

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

```
MDIFF(1)=2.2;
MDIFF(2)=0.0;
MDIFF(3)=1.0;
MDIFF(4)=2.0;
MDIFF(5)=-0.5;

MDIFF(6)=2.2;
MDIFF(7)=0.0;
MDIFF(8)=1.0;
MDIFF(9)=2.0;
MDIFF(10)=-0.5;

MDIFF(11)=2.2;
MDIFF(12)=0.0;
MDIFF(13)=1.0;
MDIFF(14)=2.0;
MDIFF(15)=-0.5;

MDIFF(16)=2.2;
MDIFF(17)=0.0;
MDIFF(18)=1.0;
MDIFF(19)=2.0;
MDIFF(20)=-0.5;

MDIFF(21)=2.2;
MDIFF(22)=0.0;
MDIFF(23)=1.0;
MDIFF(24)=2.0;
MDIFF(25)=-0.5;

MDIFF(26)=2.2;
MDIFF(27)=0.0;
MDIFF(28)=1.0;
MDIFF(29)=2.0;
MDIFF(30)=-0.5;
```

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

% การกำหนดค่าอำนาจจำแนกรวมและค่าความยากง่ายรวมในแต่ละข้อของข้อสอบร่วม 30 ข้อ
% ของนักเรียนระดับที่ 2 ทำแบบสอบระดับที่ 3

MDISC(1)=1.3;
MDISC(2)=1.0;
MDISC(3)=0.8;
MDISC(4)=1.3;
MDISC(5)=0.8;

MDISC(6)=1.3;
MDISC(7)=1.0;
MDISC(8)=0.8;
MDISC(9)=1.3;
MDISC(10)=0.6;

MDISC(11)=1.3;
MDISC(12)=1.0;
MDISC(13)=0.8;
MDISC(14)=1.3;
MDISC(15)=1.2;

MDISC(16)=1.3;
MDISC(17)=1.0;
MDISC(18)=0.8;
MDISC(19)=1.3;
MDISC(20)=1.2;

MDISC(21)=1.3;
MDISC(22)=1.0;
MDISC(23)=0.8;
MDISC(24)=1.3;
MDISC(25)=1.2;

MDISC(26)=1.3;
MDISC(27)=0.8;
MDISC(28)=0.8;
MDISC(29)=1.3;
MDISC(30)=0.6;

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

```
MDIFF(1)=3.0;
MDIFF(2)=1.5;
MDIFF(3)=2.5;
MDIFF(4)=3.0;
MDIFF(5)=2.5;

MDIFF(6)=3.0;
MDIFF(7)=1.5;
MDIFF(8)=2.5;
MDIFF(9)=3.0;
MDIFF(10)=0.0;

MDIFF(11)=3.0;
MDIFF(12)=1.5;
MDIFF(13)=2.5;
MDIFF(14)=3.0;
MDIFF(15)=0.5;

MDIFF(16)=3.0;
MDIFF(17)=1.5;
MDIFF(18)=2.5;
MDIFF(19)=3.0;
MDIFF(20)=0.5;

MDIFF(21)=3.0;
MDIFF(22)=1.5;
MDIFF(23)=2.5;
MDIFF(24)=3.0;
MDIFF(25)=0.5;

MDIFF(26)=3.0;
MDIFF(27)=2.5;
MDIFF(28)=2.5;
MDIFF(29)=3.0;
MDIFF(30)=0.0;
```

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

% การกำหนดค่าอำนาจจำแนกรวมและค่าความยากง่ายรวมในแต่ละข้อของข้อสอบร่วม 30 ข้อ
% ของนักเรียนระดับที่ 3 ทำแบบสอบระดับที่ 3

MDISC(1)=1.3;
MDISC(2)=1.0;
MDISC(3)=0.8;
MDISC(4)=1.3;
MDISC(5)=0.8;

MDISC(6)=1.3;
MDISC(7)=1.0;
MDISC(8)=0.8;
MDISC(9)=1.3;
MDISC(10)=0.6;

MDISC(11)=1.3;
MDISC(12)=1.0;
MDISC(13)=0.8;
MDISC(14)=1.3;
MDISC(15)=1.2;

MDISC(16)=1.3;
MDISC(17)=1.0;
MDISC(18)=0.8;
MDISC(19)=1.3;
MDISC(20)=1.2;

MDISC(21)=1.3;
MDISC(22)=1.0;
MDISC(23)=0.8;
MDISC(24)=1.3;
MDISC(25)=1.2;

MDISC(26)=1.3;
MDISC(27)=0.8;
MDISC(28)=0.8;
MDISC(29)=1.3;

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

```
MDISC(30)=0.6;
MDIFF(1)=2.2;
MDIFF(2)=1.0;
MDIFF(3)=2.0;
MDIFF(4)=2.2;
MDIFF(5)=2.0;

MDIFF(6)=2.2;
MDIFF(7)=1.0;
MDIFF(8)=2.0;
MDIFF(9)=2.2;
MDIFF(10)=-0.5;

MDIFF(11)=2.2;
MDIFF(12)=1.0;
MDIFF(13)=2.0;
MDIFF(14)=2.2;
MDIFF(15)=0.0;

MDIFF(16)=2.2;
MDIFF(17)=1.0;
MDIFF(18)=2.0;
MDIFF(19)=2.2;
MDIFF(20)=0.0;

MDIFF(21)=2.2;
MDIFF(22)=1.0;
MDIFF(23)=2.0;
MDIFF(24)=2.2;
MDIFF(25)=0.0;

MDIFF(26)=2.2;
MDIFF(27)=2.0;
MDIFF(28)=2.0;
MDIFF(29)=2.2;
MDIFF(30)=-0.5;
```

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

% หมายเหตุ : % จากที่กล่าวมาข้างต้นเป็นการกำหนดค่าอำนาจจำแนกรวมและค่าความยากง่าย
% รวมในแต่ละข้อ ของข้อสอบร่วม 30 ข้อ ของข้อมูล 4 ชุด คือ
% 1. ข้อมูลของของนักเรียนระดับที่ 1 ทำแบบสอบระดับที่ 2
% 2. ข้อมูลของของนักเรียนระดับที่ 2 ทำแบบสอบระดับที่ 2
% 3. ข้อมูลของของนักเรียนระดับที่ 2 ทำแบบสอบระดับที่ 3
% 4. ข้อมูลของของนักเรียนระดับที่ 3 ทำแบบสอบระดับที่ 3

% ในการประมวลผลแต่ละครั้งให้เลือกเฉพาะชุดข้อมูลใดชุดข้อมูลหนึ่งเท่านั้น โดยถ้าไม่ต้องการให้
% โปรแกรมประมวลผลในโครงสร้างใด ให้ใส่เครื่องหมาย % ไว้ข้างหน้าตัวอักษรในทุกบรรทัดของ
% ข้อมูลชุดนั้น

%%

% ใช้ในกรณีที่มีการสุ่มมุม ค่าอำนาจจำแนกรวม และค่าความยากง่ายรวมให้กับข้อสอบแต่ละข้อ

```
 %   p = [ ];
 %   fname = ['permu.txt'];
 %   fid = fopen(fname);
 %   NI = fscanf(fid,'%d\n',1);
 %   a = fscanf(fid,'%g %g\n',[1 inf]);
 %   fclose(fid);
 %   p = a;

 %p = randperm(NI);

for i=1:30
   %for j=1:(NI/30)
%   BB(p(k)) = -MDISC(i)*MDIFF(i);
%   AA(p(k),1) = cos(angle(p(k),1)*pi/180)*MDISC(i);
%   AA(p(k),2) = cos(angle(p(k),2)*pi/180)*MDISC(i);
```

% คำนวณค่าความยากง่ายของข้อสอบ และค่าอำนาจจำแนกของข้อสอบในมิติที่ 1 และ 2

```
   BB(i) = -MDISC(i)*MDIFF(i);
   AA(i,1) = cos(angle(i,1)*pi/180.0)*MDISC(i);
   AA(i,2) = cos(angle(i,2)*pi/180.0)*MDISC(i);

  %end
end
```

% แสดงค่าความยากง่ายของข้อสอบ และค่าอำนาจจำแนกของข้อสอบในมิติที่ 1 และ 2 ที่คำนวณได้

```
BB(:)
AA(:,1)
AA(:,2)
```

% จำลองค่าโอกาสการเดาของข้อสอบจำนวน 30 ข้อ จากการแจกแจงแบบ $\beta(2,50)$

```
   CC = betarnd(2,50,NI,1);
```

% อ่านแฟ้มข้อมูลโอกาสการเดาของข้อสอบจาก INPUT FILE (ใช้ในกรณีที่นำเข้าจากแฟ้มข้อมูลอื่น)

```
   % CC = [];
   % fname = ['CC_2.txt'];
   % fid = fopen(fname);
   % NN = fscanf(fid,'%d\n',1);
   % CC = fscanf(fid,'%g\n',[1 inf])
   % fclose(fid);
```

% สร้างแบบแผนการตอบข้อสอบของผู้สอบให้มีขนาด 2,000 คน จากข้อสอบร่วม 30 ข้อ

% คำนวณค่าความน่าจะเป็นในการตอบของผู้สอบ

```
% i : item, j:man
for j=1:NM
   for i=1:NI
     temp = AA(i,1)*theta1(j) + AA(i,2)*theta2(j) + BB(i);
     PP(j,i) = CC(i) +(1.0-CC(i))*exp(temp)/(1+exp(temp));
   end
end
```

```
% เปรียบเทียบความน่าจะเป็นในการตอบข้อสอบจากค่าที่คำนวณได้ กับค่าที่สุ่ม
% ถ้ามากกว่าจะให้คำตอบเป็น 1 ถ้าน้อยกว่าจะให้คำตอบเป็น 0 โดยการทำซ้ำ 20 ครั้ง
% ทำให้ได้ชุดแบบแผนการตอบของผู้สอบจำนวน 20 ชุด

for k=1:20
k

PPU = rand(NM,NI);

for j=1:NM
    for i=1:NI
     if(PP(j,i)>PPU(j,i))
         data(j,i)=1;
     else
         data(j,i)=0;
     end
    end
end


     fname = ['data_',num2str(k),'.txt'];

     fid = fopen(fname,'w');
     fprintf(fid,'%d\n',NM);
     fprintf(fid,'%d\n',NI);
     fprintf(fid,'%d\n',data);
     fclose(fid);

end

%check data
    fid = fopen(fname);
    NMI = fscanf(fid,'%d\n',1);
    NNI = fscanf(fid,'%d\n',1);
    for i=1:NNI
        for j=1:NMI
            temp = fscanf(fid,'%d\n',1);
            dataI(j,i) = temp;
        end
    end
```

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

```
if (NM == NMI) fprintf('NM is OK\n'); end
if (NI == NNI) fprintf('NN is OK\n'); end
k = 0;
   for j=1:NMI
      for i=1:NNI
         if (data(j,i) == datal(j,i)) k = k + 1; end
      end
   end
   fprintf('%d\n',k);
   if (k == NM*NI) fprintf('Data is OK'); end
%s=input('Do you want to stop - if so enter 0')
%end
```

% จัดเก็บแฟ้มข้อมูลค่าความสามารถที่แท้จริงของผู้สอบในมิติที่ 1 และ มิติที่ 2

```
    fname = ['theta_2.txt'];
    fid = fopen(fname,'w');
    fprintf(fid,'%d\n',NM);
    for i=1:NM
       fprintf(fid,'%8.4f %8.4f\n',theta1(i),theta2(i));
    end
    fclose(fid);
```

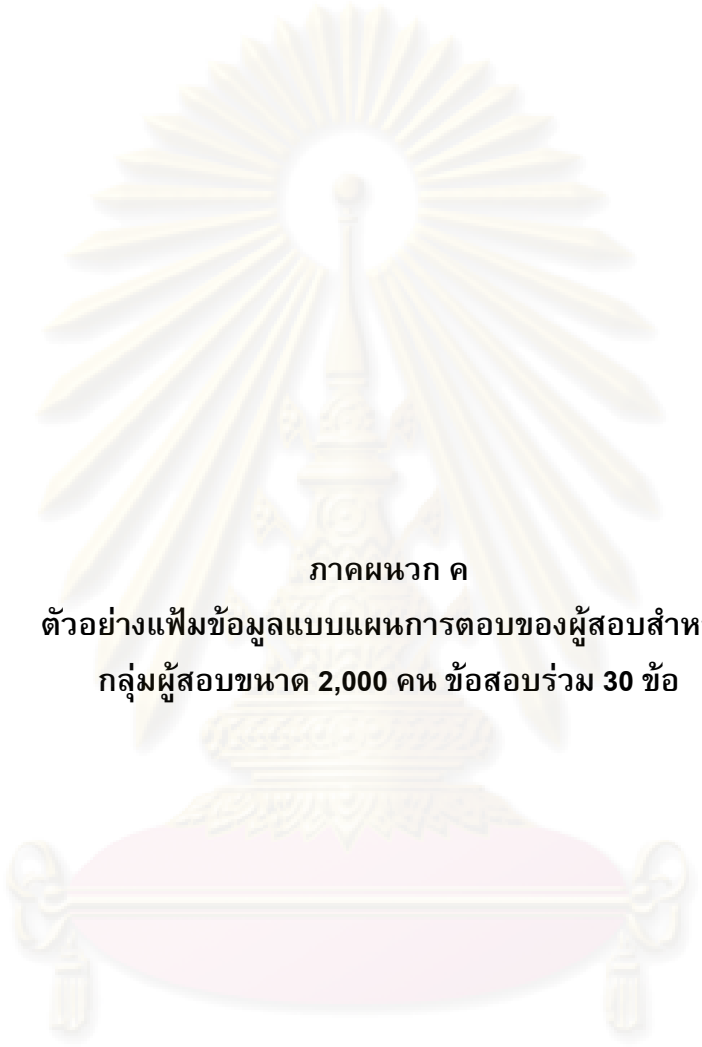
% จัดเก็บแฟ้มข้อมูลขนาดของมุมในมิติที่ 1 และ มิติที่ 2 ของข้อสอบแต่ละข้อ

```
    fname = ['angle_2.txt'];
    fid = fopen(fname,'w');
    fprintf(fid,'%d\n',NI);
    for i=1:NI
       fprintf(fid,'%8.4f %8.4f\n',angle(i,1),angle(i,2));
    end
    fclose(fid);
```

% จัดเก็บแฟ้มข้อมูลค่าโอกาสการเดาของข้อสอบในแต่ละข้อ

```
    fname = ['CC_2.txt'];
       fid = fopen(fname,'w');
       fprintf(fid,'%d\n',NI);
       for i=1:NI
          fprintf(fid,'%8.4f \n',CC(i));
       end
       fclose(fid);
```

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



ภาคผนวก ค

ตัวอย่างแฟ้มข้อมูลแบบแผนการตอบของผู้สอบสำหรับ

กลุ่มผู้สอบขนาด 2,000 คน ข้อสอบร่วม 30 ข้อ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตัวอย่างแฟ้มข้อมูลแบบแผนการตอบของผู้สอบ
สำหรับกลุ่มผู้สอบขนาด 2,000 คน ข้อสอบร่วม 30 ข้อ ของนักเรียนระดับที่ 1 ทำแบบสอบระดับที่ 2
สำหรับโครงสร้าง APSS ที่ไม่มีความสัมพันธ์ระหว่างมิติความสามารถ $(\rho = 0)$

(File “DATA_1.DAT”)

1 01001011000010001111111110001111
2 00000100000000010000100000000000
3 00100000010010100101000000000001
4 01001001000000000000000100010010
5 01000000010111101111000010 1001
6 10000001000010001000010 0100010
7 00000000010000011111111011 1101
8 00110010010000101101011111 0111
9 00111010000101011101011110 1001
10 01001000010000001100001010 0001
11 10010000101000100000010000 0000
12 00000111110000100001000100 0011
13 01000111010111100000010001 0100
14 00001010010111001110000100 1100
15 00001000100000010000100010 0011
16 01000000110110001000000110 1100

1991 01000000010000001001001010 0000
1992 00001100000100101000010010 0000
1993 10000010100000010001001000 0001
1994 01101010010100100010010010 0100
1995 10011000010000100000011010 0000
1996 00000011010010001001000100 0001
1997 01101010110100001000010000 1010
1998 01001000000101000001000110 0011
1999 00000000100010100001000010 0000
2000 00000001000101100000010000 1011

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ง
ตัวอย่างคำสั่ง MATLAB สำหรับคำนวณวิธีการเชื่อมโยงคะแนน

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

%ตัวอย่างคำสั่ง MATLAB สำหรับคำนวณวิธีการเชื่อมโยงคะแนนวิธี M และ วิธี NOP

% INPUT ค่าพารามิเตอร์ข้อสอบที่ Calibrate ได้จากโปรแกรม TESTFACT
%จากแบบสอบฐานของข้อสอบร่วม 30 ข้อ

```
base(1,1)=1;
base(2,1)=2;
base(3,1)=3;
base(4,1)=4;
base(5,1)=5;
base(6,1)=6;
base(7,1)=7;
base(8,1)=8;
base(9,1)=9;
base(10,1)=10;
base(11,1)=11;
base(12,1)=12;
base(13,1)=13;
base(14,1)=14;
base(15,1)=15;
base(16,1)=16;
base(17,1)=17;
base(18,1)=18;
base(19,1)=19;
base(20,1)=20;
base(21,1)=21;
base(22,1)=22;
base(23,1)=23;
base(24,1)=24;
base(25,1)=25;
base(26,1)=26;
base(27,1)=27;
base(28,1)=28;
base(29,1)=29;
base(30,1)=30;
```

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

% INPUT ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบในมิติที่ 1 (a_1) จากแบบสอบฐาน

```
base(1,2)=0.743;
base(2,2)=1.140;
base(3,2)=1.044;
base(4,2)=0.759;
base(5,2)=0.489;
base(6,2)=0.900;
base(7,2)=0.843;
base(8,2)=0.844;
base(9,2)=0.700;
base(10,2)=0.521;
base(11,2)=0.925;
base(12,2)=0.988;
base(13,2)=0.804;
base(14,2)=0.696;
base(15,2)=0.658;
base(16,2)=0.152;
base(17,2)=0.238;
base(18,2)=0.271;
base(19,2)=0.138;
base(20,2)=0.069;
base(21,2)=0.126;
base(22,2)=0.208;
base(23,2)=0.107;
base(24,2)=0.154;
base(25,2)=0.109;
base(26,2)=0.026;
base(27,2)=0.022;
base(28,2)=0.074;
base(29,2)=0.041;
base(30,2)=0.005;
```

% INPUT ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบในมิติที่ 2 (a_2) จากแบบสอบฐาน

```
base(1,3)=0.145;
base(2,3)=-0.068;
base(3,3)=0.156;
base(4,3)=-0.002;
base(5,3)=0.136;
```

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

```
base(6,3)=0.184;
base(7,3)=0.085;
base(8,3)=0.089;
base(9,3)=0.143;
base(10,3)=0.132;
base(11,3)=0.165;
base(12,3)=0.205;
base(13,3)=0.278;
base(14,3)=0.069;
base(15,3)=0.258;
base(16,3)=0.795;
base(17,3)=0.832;
base(18,3)=0.735;
base(19,3)=0.676;
base(20,3)=0.456;
base(21,3)=0.804;
base(22,3)=0.857;
base(23,3)=0.778;
base(24,3)=0.594;
base(25,3)=0.433;
base(26,3)=0.997;
base(27,3)=0.871;
base(28,3)=0.758;
base(29,3)=0.622;
base(30,3)=0.696;

% INPUT ค่าพารามิเตอร์ความยากง่ายของข้อสอบจากแบบสอบฐาน
base(1,4)=-1.684;
base(2,4)=-0.553;
base(3,4)=-1.357;
base(4,4)=-1.886;
base(5,4)=-0.276;
base(6,4)=-1.959;
base(7,4)=-0.195;
base(8,4)=-0.874;
base(9,4)=-1.641;
base(10,4)=-0.275;
base(11,4)=-2.006;
```

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

```
base(12,4)=-0.258;
base(13,4)=-0.885;
base(14,4)=-1.404;
base(15,4)=-0.786;
base(16,4)=-1.531;
base(17,4)=0.021;
base(18,4)=-0.635;
base(19,4)=-1.138;
base(20,4)=-0.063;
base(21,4)=-1.519;
base(22,4)=-0.017;
base(23,4)=-0.668;
base(24,4)=-1.106;
base(25,4)=-0.246;
base(26,4)=-1.676;
base(27,4)=-0.035;
base(28,4)=-0.786;
base(29,4)=-1.247;
base(30,4)=-0.338;

% INPUT ค่าพารามิเตอร์ข้อสอบที่ Calibrate ได้จากโปรแกรม TESTFACT
%จากแบบสอบปรับเทียบคะแนนของข้อสอบร่วม 30 ข้อ

equate(1,1)=1;
equate(2,1)=2;
equate(3,1)=3;
equate(4,1)=4;
equate(5,1)=5;
equate(6,1)=6;
equate(7,1)=7;
equate(8,1)=8;
equate(9,1)=9;
equate(10,1)=10;
equate(11,1)=11;
equate(12,1)=12;
equate(13,1)=13;
equate(14,1)=14;
equate(15,1)=15;
```

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

```
equate(16,1)=16;
equate(17,1)=17;
equate(18,1)=18;
equate(19,1)=19;
equate(20,1)=20;
equate(21,1)=21;
equate(22,1)=22;
equate(23,1)=23;
equate(24,1)=24;
equate(25,1)=25;
equate(26,1)=26;
equate(27,1)=27;
equate(28,1)=28;
equate(29,1)=29;
equate(30,1)=30;

% INPUT ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบในมิติที่ 1 (a1) จากแบบสอบปรับเทียบคะแนน
equate(1,2)=0.683;
equate(2,2)=0.869;
equate(3,2)=0.704;
equate(4,2)=0.661;
equate(5,2)=0.606;
equate(6,2)=0.807;
equate(7,2)=0.859;
equate(8,2)=0.793;
equate(9,2)=0.617;
equate(10,2)=0.439;
equate(11,2)=0.937;
equate(12,2)=0.807;
equate(13,2)=0.691;
equate(14,2)=0.670;
equate(15,2)=0.561;
equate(16,2)=0.415;
equate(17,2)=0.443;
equate(18,2)=0.197;
equate(19,2)=0.177;
equate(20,2)=0.222;
equate(21,2)=0.158;
```

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

equate(22,2)=0.178;
equate(23,2)=0.225;
equate(24,2)=0.189;
equate(25,2)=0.057;
equate(26,2)=-0.040;
equate(27,2)=0.353;
equate(28,2)=0.159;
equate(29,2)=-0.101;
equate(30,2)=0.175;

% INPUT ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบในมิติที่ 2 (a_2) จากแบบสอบปรับเทียบคะแนน
equate(1,3)=0.156;
equate(2,3)=0.017;
equate(3,3)=-0.024;
equate(4,3)=-0.020;
equate(5,3)=0.115;
equate(6,3)=0.340;
equate(7,3)=0.167;
equate(8,3)=0.159;
equate(9,3)=0.173;
equate(10,3)=0.138;
equate(11,3)=0.253;
equate(12,3)=0.269;
equate(13,3)=0.140;
equate(14,3)=0.233;
equate(15,3)=0.078;
equate(16,3)=0.896;
equate(17,3)=1.280;
equate(18,3)=0.917;
equate(19,3)=0.727;
equate(20,3)=0.869;
equate(21,3)=0.752;
equate(22,3)=1.299;
equate(23,3)=0.933;
equate(24,3)=0.680;
equate(25,3)=0.745;
equate(26,3)=1.073;
equate(27,3)=1.588;

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

```
equate(28,3)=1.026;
equate(29,3)=0.913;
equate(30,3)=0.621;

% INPUT ค่าพารามิเตอร์ความยากง่ายของข้อสอบจากแบบสอบปรับเทียบคะแนน
equate(1,4)=-2.224;
equate(2,4)=-0.721;
equate(3,4)=-1.252;
equate(4,4)=-1.764;
equate(5,4)=-0.642;
equate(6,4)=-2.629;
equate(7,4)=-0.756;
equate(8,4)=-1.457;
equate(9,4)=-1.613;
equate(10,4)=-0.599;
equate(11,4)=-2.815;
equate(12,4)=-0.810;
equate(13,4)=-1.308;
equate(14,4)=-1.695;
equate(15,4)=-0.445;
equate(16,4)=-3.398;
equate(17,4)=-1.980;
equate(18,4)=-1.858;
equate(19,4)=-2.165;
equate(20,4)=-1.642;
equate(21,4)=-2.626;
equate(22,4)=-1.799;
equate(23,4)=-2.097;
equate(24,4)=-2.178;
equate(25,4)=-1.154;
equate(26,4)=-3.418;
equate(27,4)=-2.508;
equate(28,4)=-2.276;
equate(29,4)=-2.618;
equate(30,4)=-1.138;
```

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

```
% กำหนดชื่อตัวแปรสำหรับแบบสอบฐาน
%base=x;
global A1 A2 d A1o A2o do D;
d=base(:,4);
A1=base(:,2);
A2=base(:,3);
BASE=base(:,2:3);

% กำหนดชื่อตัวแปรสำหรับแบบสอบปรับเทียบคะแนน
%equate=y;
do=equate(:,4);
TA1=equate(:,2);
TA2=equate(:,3);
EQUATE_1=equate(:,2:3);
```

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

%ตัวอย่างคำสั่ง MATLAB สำหรับคำนวณวิธีการเชื่อมโยงคะแนนโดยวิธี M

```
% การแปลงคะแนนค่าพารามิเตอร์อำนาจจำแนกข้อสอบจากแบบสอบปรับเทียบคะแนนไปยังแบบสอบฐาน
% a1,a2_M Method
dM=[d do];
cordM=corrcoef(dM);
BASEEQUATE_1=[BASE EQUATE_1];
COR_obs=corrcoef(BASEEQUATE_1);

disp('Orthogonal Procrustes Rotation, Schoneman, 1996');
S=EQUATE_1'*BASE;
STS=S'*S;
SST=S*S';
[U,S,V]=svd(STS)
V1=U;
D1=S;
V1=V;
[U,S,V]=svd(SST);
W1=U;
D1=S;
W1=V;
ESTRM_1=W1*V1';
ESTRM=ESTRM_1;
EQUATErot=EQUATE_1*ESTRM_1;

disp('Rotation Matrix, T');
ESTRM;
T = ESTRM
EQUATErot;
A1o=EQUATErot(:,1);
A2o=EQUATErot(:,2);

disp ('Diagonal Dilation Matrix, K');
EQUATE_1_C=EQUATE_1;
BASE_C=BASE;
LEFT=ESTRM'*EQUATE_1_C'*EQUATE_1_C*ESTRM;
RIGHT=BASE_C'*EQUATE_1_C*ESTRM;
DEN=inv(diag(diag(LEFT)));
NUM=diag(diag(RIGHT));
```

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

```
K=NUM*DEN;
K

TK = T*K
% Finf a*

EQUATE_2=equate(:,2:3);
BASE_2=base(:,2:3);

for i=1:30
    %EQUATE_2(i,1:2)*TK
    ai_(i,:) = EQUATE_2(i,1:2)*TK;
end

% การคำนวณดัชนีบ่งชี้คุณภาพของการเชื่อมโยงคะแนน (RMSE) ของ a1 และ a2
ai_(:,1:2)

sum2 = 0.0;

for i=1:30
   sum2 = sum2 + (ai_(i,1)-BASE_2(i,1))^2;
end

RMSE1 = sqrt(sum2/29)
RMSE2 = sqrt(sum2/29)



% การแปลงคะแนนค่าพารามิเตอร์ความยากง่ายของข้อสอบจากแบบสอบปรับเทียบคะแนนไปยังแบบสอบฐาน
% d_M Method

for i=1:30
AAA(i,1)= equate(i,2)*T(1,1)+equate(i,3)*T(2,1);
AAA(i,2)= equate(i,2)*T(1,2)+equate(i,3)*T(2,2);
Dcb(i)= equate(i,4)-base(i,4);
end
MM = Dcb*AAA*inv(AAA'*AAA)

for i=1:30
   di_(i)= equate(i,4) - MM(1)*(equate(i,2)*T(1,1)+equate(i,3)*T
(2,1)) -MM(2)*(equate(i,2)*T(1,2)+equate(i,3)*T(2,2));
End
```

```
% การคำนวณดัชนีบ่งชี้คุณภาพของการเชื่อมโยงคะแนน (RMSE) ของ d
di_(:)
base(:,4)

sum2 = 0.0;

for i=1:30
   sum2 = sum2 + (di_(i)-base(i,4))^2;
end

RMSE_di = sqrt(sum2/29)
```

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

%ตัวอย่างคำสั่ง MATLAB สำหรับคำนวณวิธีการเชื่อมโยงคะแนนโดยวิธี NOP

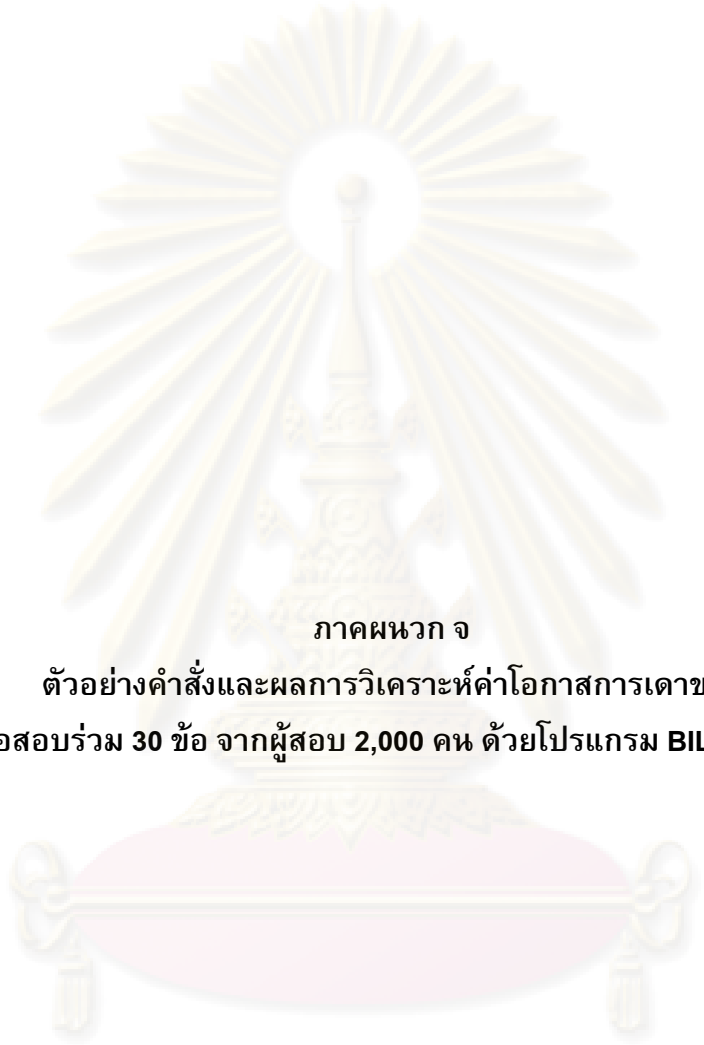
```
% การแปลงคะแนนค่าพารามิเตอร์อำนาจจำแนกข้อสอบจากแบบสอบปรับเทียบคะแนนไปยังแบบสอบฐาน
% a1,a2-NOP_METHOD

T1 = inv(EQUATE_1'*EQUATE_1)
T2 = EQUATE_1'*BASE

T = T1*T2



% Finf a*

EQUATE_2=equate(:,2:3);
BASE_2=base(:,2:3);

for i=1:30
    %EQUATE_2(i,1:2)*T
    ai_(i,:) = EQUATE_2(i,1:2)*T;
end
T

% การคำนวณดัชนีบ่งชี้คุณภาพของการเชื่อมโยงคะแนน (RMSE) ของ a1 และ a2
ai_(:,1:2)

sum2 = 0.0;

for i=1:30
   sum2 = sum2 + (ai_(i,1)-BASE_2(i,1))^2;
end

RMSE1 = sqrt(sum2/29)
RMSE2 = sqrt(sum2/29)

% การแปลงคะแนนค่าพารามิเตอร์ความยากง่ายของข้อสอบจากแบบสอบปรับเทียบคะแนนไปยังแบบสอบฐาน
% d_NOP Method

for i=1:30
AAA(i,1)= equate(i,2)*T(1,1)+equate(i,3)*T(2,1);
```

```
AAA(i,2)= equate(i,2)*T(1,2)+equate(i,3)*T(2,2);
Dcb(i)= equate(i,4)-base(i,4);
end
MM = -Dcb*AAA*inv(AAA'*AAA)

for i=1:30
   di_(i)= equate(i,4) +
MM(1)*(equate(i,2)*T(1,1)+equate(i,3)*T(2,1))+MM(2)*(equate(i,2)*T(1,2)+equate(i,3)*T(2,2));
end

% การคำนวณดัชนีบ่งชี้คุณภาพของการเชื่อมโยงคะแนน (RMSE) ของ d
di_(:)
base(:,4)

sum2 = 0.0;

for i=1:30
   sum2 = sum2 + (di_(i)-base(i,4))^2;
end

RMSE_di = sqrt(sum2/29)
```

% หมายเหตุ:

```
% (1) สำหรับการคำนวณดัชนี BIAS และสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าพารามิเตอร์
% จากแบบสอบฐานและค่าพารามิเตอร์ที่ได้จากการแปลงคะแนนจากแบบสอบปรับเทียบ
% คะแนน ผู้วิจัยวิเคราะห์โดยใช้โปรแกรม EXCEL และ SPSS for Windows Version
% 14.0 ตามลำดับ
% (2) ในการประมวลผลแต่ละครั้งให้เลือกเฉพาะวิธีการเชื่อมโยงคะแนนเพียงวิธีเดียว
% เท่านั้น ถ้าไม่ต้องการให้โปรแกรมประมวลผลในวิธีการใด ให้ใส่เครื่องหมาย % ไว้
% ข้างหน้าตัวอักษรในทุกบรรทัดของคำสั่งวิธีการนั้น
```

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



ภาคผนวก จ

ตัวอย่างคำสั่งและผลการวิเคราะห์ค่าโอกาสการเดาของ

ข้อสอบร่วม 30 ข้อ จากผู้สอบ 2,000 คน ด้วยโปรแกรม BILOG-MG

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

การวิเคราะห์ด้วยโปรแกรม BILOG-MG for Windows

คุณสมบัติของโปรแกรม

1. ใช้ประมาณค่าพารามิเตอร์ของข้อสอบและผู้สอบ (a,b,c และ θ) ภายใต้ ทฤษฎี IRT โมเดล 1,2 และ 3 พารามิเตอร์

2. เหมาะสำหรับแบบแผนการตอบข้อสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค

ตัวอย่าง File คำสั่ง เพื่อประมาณค่าโอกาสการเดาของข้อสอบ (C)

```
DATA_1.BLM - TRADITIONAL IRT ANALYSIS OF A THIRTY-ITEM TEST
FROM A TWO-STAGE TEST OF MATHEMATICS AT THE FIRST-GRADE LEVEL
>COMMENTS
L3 STUDENT TAKE L3 TEST
>GLOBAL  DFNAME='DATA_1.DAT', NPARM=3, SAVE;
>SAVE    PARM='DATA_1.PAR', SCORE='DATA_1.SCO';
>LENGTH  NITEMS=30;
>INPUT   NTOTAL=30, NALT=3, NIDCHAR=4,
         KFNAME='DATA_1.KEY', OFNAME='DATA_1.OMT';
>ITEMS   INAMES=(ITEM01(1)ITEM30);
>TEST1   TNAME='TEST_L12', INUMBER=(1(1)30);
(4A1,1X,30A1)
>CALIB   NQPT=31, CYCLES=25, NEWTON=10, CRIT=0.001,
ACCEL=0.0, CHI=15, PLOT=1;
>SCORE   NOPRINT, RSCTYPE=4, INFO=2, POP;
```

เมื่อ	DFNAME='DATA_1.DAT'	หมายถึง	File ข้อมูลดิบ
	KFNAME='DATA_1.KEY'	หมายถึง	File คำตอบ(กำหนดให้มีค่าเท่ากับ 1)
	OFNAME='DATA_1.OMT'	หมายถึง	File ของค่า Missing (กำหนดให้มีค่าเท่ากับ 9)
	NPARM =3	หมายถึง	จำนวนพารามิเตอร์
	NALT=3	หมายถึง	จำนวนตัวเลือก มี 3 ค่า คือ 0, 1 และ 9
	NITEMS=30	หมายถึง	จำนวนข้อสอบ
	NIDCHAR=4	หมายถึง	จำนวน ID
	4A1	หมายถึง	จำนวน ID
	1X	หมายถึง	ให้เว้นไป 1 คอลัมภ์
	30A1	หมายถึง	จำนวนข้อสอบที่วิเคราะห์
	INF=2	หมายถึง	ให้ระบุรูปแบบของสารสนเทศของแบบสอบโดยรายงานเป็นโค้งสารสนเทศของแบบสอบและตารางค่าสารสนเทศของแบบสอบ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

Output จากการวิเคราะห์โดยใช้โปรแกรม BILOG-MG

ผลลัพธ์ที่ได้จากการประมาณค่าพารามิเตอร์ของข้อสอบและผู้สอบ (a,b,c และ θ) ภายใต้ ทฤษฎี IRT โมเดล 1,2 และ 3 พารามิเตอร์ ของโปรแกรม BILOG-MG มีทั้งหมด 3 Phase โดยมีรายละเอียด ดังนี้

Phase 1 (INPUT) รายงานค่าที่ INPUT เข้าไปในโปรแกรม และรายงานค่าสถิติเบื้องต้น ประกอบด้วย Omits and Attempts, Items and Forms, Multiple Groups, Subtests, Case Weights, Samples, Classical Item Statistics

Phase 2 (CALIBRATE) รายงานค่าพารามิเตอร์ของข้อสอบที่ประมาณได้ ประกอบด้วย ค่าอำนาจจำแนกของข้อสอบ (a) ค่าความยากง่ายของข้อสอบ (b) และค่าโอกาสการเดาของข้อสอบ (c)

Phase 3 (SCORE) รายงานผลการตอบข้อสอบเป็นรายบุคคล ประกอบด้วย ค่าความสามารถที่แท้จริงของผู้สอบ (θ) คะแนนที่สังเกตได้ของผู้สอบ (Observe Scores) รวมถึงรายงานค่าฟังก์สารสนเทศของข้อสอบแต่ละข้อและฟังก์ชันสารสนเทศของแบบสอบทั้งฉบับโดยรายงานเป็นโค้งสารสนเทศของแบบสอบและตารางค่าสารสนเทศของแบบสอบ

ในที่นี้ผู้วิจัยได้นำเสนอ Printout เฉพาะใน Phase 2 ที่แสดงค่าโอกาสการเดาของข้อสอบ (C) เพื่อนำผลการวิเคราะห์ที่ได้เข้าสู่โปรแกรม TESTFACT ในลำดับต่อไป

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตัวอย่าง Printout แสดงค่าโอกาสการเดาของข้อสอบ (C) ใน Phase 2

BILOG-MG V3.0

REV 19990329.1300

BILOG-MG ITEM MAINTENANCE PROGRAM: LOGISTIC ITEM RESPONSE MODEL

*** BILOG-MG ITEM MAINTENANCE PROGRAM ***

*** PHASE 2 ***

DATA_1.BLM - TRADITIONAL IRT ANALYSIS OF A THIRTY-ITEM TEST FROM
A TWO-STAGE TEST OF MATHEMATICS AT THE FIRST-GRADE LEVEL

>CALIB NQPT=31, CYCLES=25, NEWTON=10, CRIT=0.001, ACCEL=0.0, CHI=15, PLOT=1;

CALIBRATION PARAMETERS
======================
MAXIMUM NUMBER OF EM CYCLES: 25

MAXIMUM NUMBER OF NEWTON CYCLES: 10
CONVERGENCE CRITERION: 0.0010
ACCELERATION CONSTANT: 0.0000

LATENT DISTRIBUTION: NORMAL PRIOR FOR EACH GROUP
PLOT EMPIRICAL VS. FITTED ICC'S: YES, FOR ITEMS WITH FIT PROBABILITY
LESS THAN 1.00000
DATA HANDLING: DATA ON SCRATCH FILE
CONSTRAINT DISTRIBUTION ON ASYMPTOTES: YES
CONSTRAINT DISTRIBUTION ON SLOPES: YES
CONSTRAINT DISTRIBUTION ON THRESHOLDS: NO
SOURCE OF ITEM CONSTRAINT DISTIBUTION
MEANS AND STANDARD DEVIATIONS: PROGRAM DEFAULTS

****

CALIBRATION OF MAINTEST

TEST_L12

****

CONSTRAINT DISTRIBUTIONS ON ITEM PARAMETERS
(THRESHOLDS, NORMAL; SLOPES, LOG-NORMAL; GUESSING, BETA)

	THRESHOLDS		SLOPES		ASYMPTOTES	
ITEM	MU	SIGMA	MU	SIGMA	ALPHA	BETA
ITEM01	-	-	1.000	1.649	7.67	14.33
ITEM02	-	-	1.000	1.649	7.67	14.33
ITEM03	-	-	1.000	1.649	7.67	14.33
⋮			⋮	⋮	⋮	⋮
ITEM28	-	-	1.000	1.649	7.67	14.33
ITEM29	-	-	1.000	1.649	7.67	14.33
ITEM30	-	-	1.000	1.649	7.67	14.33

[E-M CYCLES]

-2 LOG LIKELIHOOD = 72446.957
CYCLE 1; LARGEST CHANGE= 2.65818
-2 LOG LIKELIHOOD = 63685.354
CYCLE 2; LARGEST CHANGE= 0.70210
-2 LOG LIKELIHOOD = 63447.904
⋮ ⋮
CYCLE 24; LARGEST CHANGE= 0.01926
-2 LOG LIKELIHOOD = 63352.788
CYCLE 25; LARGEST CHANGE= 0.01794

====> NOTE: CONVERGENCE HAS NOT BEEN REACHED TO CRITERION = 0.00100

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

```
 [NEWTON CYCLES]

 -2 LOG LIKELIHOOD:         63352.5277
 CYCLE    26;   LARGEST CHANGE=   0.04999
 -2 LOG LIKELIHOOD:         63352.3242
 CYCLE    27;   LARGEST CHANGE=   0.02604
    .                          .
    .                          .
-2 LOG LIKELIHOOD:         63351.2560
 CYCLE    34;   LARGEST CHANGE=   0.00872
 -2 LOG LIKELIHOOD:         63351.1982
 CYCLE    35;   LARGEST CHANGE=   0.00729

 ====> NOTE: CONVERGENCE HAS NOT BEEN REACHED TO CRITERION =  0.00100


 INTERVAL COUNTS FOR COMPUTATION OF ITEM CHI-SQUARES
  ------------------------------------------------------------------------------
         0.    29.   292.   403.   373.   293.   254.   180.   176.
  ------------------------------------------------------------------------------

 INTERVAL AVERAGE THETAS
  ------------------------------------------------------------------------------
     ******* -1.608 -1.272 -0.786 -0.264  0.240  0.757  1.263  1.950
  ------------------------------------------------------------------------------
1
```

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

SUBTEST TEST_L12 ; ITEM PARAMETERS AFTER CYCLE 35

ITEM	INTERCEPT S.E.	SLOPE S.E.	THRESHOLD S.E.	LOADING S.E.	**ASYMPTOTE** S.E.	CHISQ (PROB)	DF
ITEM01	-2.174	0.753	2.888	0.601	0.091	7.3 (0.3981)	7.0
	0.411*	0.223*	0.399*	0.179*	0.016*		
ITEM02	-0.523	0.650	0.804	0.545	0.099	118.0 (0.0000)	7.0
	0.106*	0.074*	0.110*	0.062*	0.034*		
ITEM03	-1.054	0.542	1.946	0.476	0.112	32.8 (0.0000)	7.0
	0.187*	0.105*	0.185*	0.092*	0.033*		
ITEM04	-1.636	0.590	2.775	0.508	0.112	25.9 (0.0011)	8.0
	0.302*	0.160*	0.359*	0.137*	0.024*		
ITEM05	-0.506	0.582	0.870	0.503	0.263	31.4 (0.0001)	7.0
	0.214*	0.118*	0.228*	0.102*	0.064*		
ITEM06	-2.481	0.948	2.616	0.688	0.067	4.5 (0.7225)	7.0
	0.438*	0.251*	0.290*	0.182*	0.011*		
ITEM07	-0.550	0.771	0.713	0.611	0.114	74.6 (0.0000)	7.0
	0.127*	0.096*	0.104*	0.076*	0.037*		
ITEM08	-1.229	0.732	1.679	0.591	0.165	18.4 (0.0104)	7.0
	0.253*	0.155*	0.133*	0.125*	0.032*		
ITEM09	-1.473	0.657	2.243	0.549	0.071	18.0 (0.0118)	7.0
	0.207*	0.125*	0.198*	0.105*	0.020*		
ITEM10	-0.494	0.487	1.014	0.438	0.255	36.6 (0.0000)	7.0
	0.208*	0.101*	0.270*	0.090*	0.067*		
ITEM11	-2.486	0.894	2.782	0.666	0.057	10.5 (0.1636)	7.0
	0.419*	0.238*	0.336*	0.177*	0.010*		

ITEM12	-0.605	0.835	0.725	0.641	0.087	104.7 7.0
	0.111*	0.093*	0.083*	0.071*	0.029*	(0.0000)
ITEM13	-1.131	0.670	1.688	0.557	0.095	41.9 7.0
	0.179*	0.114*	0.129*	0.094*	0.027*	(0.0000)
ITEM14	-1.548	0.751	2.060	0.601	0.072	24.1 7.0
	0.225*	0.144*	0.167*	0.115*	0.019*	(0.0011)
ITEM15	-0.330	0.518	0.636	0.460	0.183	73.0 7.0
	0.144*	0.077*	0.212*	0.068*	0.059*	(0.0000)
ITEM16	-3.209	1.151	2.788	0.755	0.053	9.8 7.0
	0.692*	0.358*	0.329*	0.235*	0.008*	(0.2012)
ITEM17	-1.548	1.304	1.187	0.793	0.295	15.7 8.0
	0.376*	0.286*	0.082*	0.174*	0.025*	(0.0464)
ITEM18	-1.457	0.829	1.757	0.638	0.144	11.5 7.0
	0.275*	0.172*	0.125*	0.132*	0.025*	(0.1200)
ITEM19	-2.020	0.825	2.448	0.636	0.122	10.7 8.0
	0.392*	0.220*	0.258*	0.170*	0.018*	(0.2207)
ITEM20	-1.344	0.863	1.558	0.653	0.420	3.9 8.0
	0.392*	0.238*	0.155*	0.180*	0.031*	(0.8663)
ITEM21	-2.639	0.931	2.836	0.681	0.055	8.8 7.0
	0.481*	0.264*	0.354*	0.193*	0.010*	(0.2700)
ITEM22	-1.144	0.908	1.259	0.672	0.236	24.8 7.0
	0.276*	0.188*	0.107*	0.139*	0.034*	(0.0008)
ITEM23	-1.756	0.921	1.907	0.677	0.152	7.5 7.0
	0.344*	0.212*	0.141*	0.156*	0.021*	(0.3748)
ITEM24	-2.116	0.844	2.508	0.645	0.134	2.8 7.0
	0.431*	0.240*	0.283*	0.184*	0.017*	(0.9070)

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ITEM25	-0.832	0.550	1.511	0.482	0.376	32.6 8.0
	0.302*	0.152*	0.251*	0.133*	0.053*	(0.0001)
ITEM26	-2.876	0.894	3.216	0.667	0.078	8.0 8.0
	0.621*	0.315*	0.546*	0.235*	0.010*	(0.4340)
ITEM27	-2.030	1.553	1.307	0.841	0.326	6.5 8.0
	0.516*	0.381*	0.077*	0.206*	0.020*	(0.5868)
ITEM28	-2.071	1.077	1.923	0.733	0.212	7.5 8.0
	0.451*	0.273*	0.144*	0.186*	0.019*	(0.4879)
ITEM29	-2.689	1.001	2.686	0.707	0.128	4.4 7.0
	0.604*	0.321*	0.334*	0.227*	0.013*	(0.7368)
ITEM30	-0.903	0.597	1.513	0.513	0.365	18.8 8.0
	0.301*	0.157*	0.220*	0.135*	0.048*	(0.0159)

---

* STANDARD ERROR

LARGEST CHANGE = 0.007295 794.8 219.0

(0.0000)

PARAMETER	MEAN	STN DEV
ASYMPTOTE	0.165	0.105
SLOPE	0.821	0.238
LOG(SLOPE)	-0.235	0.276
THRESHOLD	1.862	0.767

78912 BYTES OF NUMERICAL WORKSPACE USED OF 8192000 AVAILABLE IN PHASE-2

4188 BYTES OF CHARACTER WORKSPACE USED OF 2048000 AVAILABLE IN PHASE-2

01/20/2008 21:41:01

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ฉ
ตัวอย่างคำสั่งและผลการวิเคราะห์เพื่อประมาณค่าพารามิเตอร์อำนาจจำแนก และค่าความยากง่ายของข้อสอบ 30 ข้อ จากผู้สอบ 2,000 คน ด้วยโปรแกรม TESTFACT

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

การวิเคราะห์ด้วยโปรแกรม TESTFACT

คุณสมบัติของโปรแกรม

1. ใช้ประมาณค่าพารามิเตอร์ของข้อสอบและผู้สอบภายใต้ทฤษฎี MIRT โมเดล 1,2 และ 3 พารามิเตอร์
2. ใช้ผลการวิเคราะห์ค่าพารามิเตอร์โอกาสการเดาของข้อสอบ (c) จากโปรแกรม BILOG-MG
3. เหมาะสำหรับแบบแผนการตอบข้อสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค

ตัวอย่าง File คำสั่งสำหรับการหมุนแกนแบบตั้งฉาก

```
>TITLE
DATA_1.TSF - L1 TAKE L2 TEST DATA ADAPTIVE FULL-INFORMATION ITEM FACTOR
               ANALYSIS;    COUNTED RESPONSE PATTERNS
>PROBLEM       NITEMS=30, RESPONSE=3; <=== OMIT response should be included
>COMMENT       FACTOR 2 FULL INFORMATION
>NAMES         ITEM1,ITEM2,ITEM3,ITEM4,ITEM5,ITEM6,ITEM7,ITEM8,ITEM9,
               ITEM10,ITEM11,ITEM12,ITEM13,ITEM14,ITEM15,ITEM16,
               ITEM17,ITEM18,ITEM19,ITEM20,ITEM21,ITEM22,ITEM23,
               ITEM24,ITEM25,ITEM26,ITEM27,ITEM28,ITEM29,ITEM30;
>RESPONSE      ' ','0','1';  <==== OMIT response comes first
>KEY           11111111111111111111111111111;
>RELIABILITY   ALPHA;
>TETRACHORIC   NDEC=3, LIST;
>FACTOR        NFAC=2, NROOT=10, NIT=5, ROTATE=VARIMAX, RESIDUAL;
>FULL          CYCLES=100, CPARMS=(0.091,0.099,0.112,0.112,
               0.263,0.067,0.114,0.165,0.071,0.255,0.057,
               0.087,0.095,0.072,0.183,0.053,0.295,0.144,
               0.122,0.420,0.055,0.236,0.152,0.134,0.376,0.078,
               0.326,0.212,0.128,0.365), OMIT=RECODE;
>TECHNICAL     PRECISION=0.0001,QUAD=10;
>SCORE         METHOD=2;
>SAVE          CORR, ROTATE, PARM;
>INPUT         NIDCHAR=4, SCORES, FILE='DATA_1.DAT';
(4A1,1x,30A1)
>STOP
```

ตัวอย่าง File คำสั่งสำหรับการหมุนแกนแบบไม่ตั้งฉาก

```
>TITLE
DATA_1.TSF - L1 TAKE L2 TEST DATA ADAPTIVE FULL-INFORMATION ITEM FACTOR
               ANALYSIS;    COUNTED RESPONSE PATTERNS
>PROBLEM       NITEMS=30, RESPONSE=3; <=== OMIT response should be included
>COMMENT       FACTOR 2 FULL INFORMATION
>NAMES         ITEM1,ITEM2,ITEM3,ITEM4,ITEM5,ITEM6,ITEM7,ITEM8,ITEM9,
               ITEM10,ITEM11,ITEM12,ITEM13,ITEM14,ITEM15,ITEM16,
               ITEM17,ITEM18,ITEM19,ITEM20,ITEM21,ITEM22,ITEM23,
               ITEM24,ITEM25,ITEM26,ITEM27,ITEM28,ITEM29,ITEM30;
>RESPONSE      ' ','0','1';  <==== OMIT response comes first
>KEY           111111111111111111111111111111;
>RELIABILITY   ALPHA;
>TETRACHORIC   NDEC=3, LIST;
>FACTOR        NFAC=2, NROOT=10, NIT=5, ROTATE=PROMAX, RESIDUAL;
>FULL          CYCLES=100, CPARMS=(0.091,0.099,0.112,0.112,
               0.263,0.067,0.114,0.165,0.071,0.255,0.057,
               0.087,0.095,0.072,0.183,0.053,0.295,0.144,
               0.122,0.420,0.055,0.236,0.152,0.134,0.376,0.078,
               0.326,0.212,0.128,0.365), OMIT=RECODE;
>TECHNICAL     PRECISION=0.0001,QUAD=10;
>SCORE         METHOD=2;
>SAVE          CORR, ROTATE, PARM;
>INPUT         NIDCHAR=4, SCORES, FILE='DATA_1.DAT';
(4A1,1x,30A1)
>STOP
```

เมื่อ FILE='DATA_1.DAT' หมายถึง File ข้อมูลดิบ

DATA_1.TSF หมายถึง File ข้อมูลคำสั่งของโปรแกรม TESTFACT

NITEMS=30 หมายถึง จำนวนข้อสอบ

RESPONSE=3 หมายถึง จำนวนตัวเลือก มี 3 ค่า คือ 0, 1 และ 9

NDEC=3 หมายถึง จำนวนหลักของเลขทศนิยม

NFAC=2 หมายถึง จำนวนมิติที่ศึกษา

ROTATE=VARIMAX หมายถึง การหมุนแกน มี 2 แบบ คือ

/VARIMAX VARIMAX (ไม่มีความสัมพันธ์ระหว่างมิติความสามารถ)

และ PROMAX (มีความสัมพันธ์ระหว่างมิติความสามารถ)

RESIDUAL หมายถึง การให้โปรแกรมคำนวณค่าส่วนที่เหลือ (Residual)

CYCLES=100 หมายถึง จำนวนรอบที่ทำการหมุน

CPARMS หมายถึง ให้ INPUT ค่าโอกาสการเดาของข้อสอบ (C)

METHOD=2 หมายถึง วิธีการที่ใช้ในการประมาณค่า มี 2 วิธี

1 = Maximum A Posteriori (MAP) estimation

2 = Bayes (EAP) estimation

NIDCHAR=4 หมายถึง จำนวน ID

4A1 หมายถึง จำนวน ID

1X หมายถึง ให้เว้นไป 1 คอลัมภ์

30A1 หมายถึง จำนวนข้อสอบที่วิเคราะห์

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตัวอย่าง Printout เพื่อประมาณค่าพารามิเตอร์ของข้อสอบในมิติที่ 1 และ 2

TESTFACT VERSION 4.0

TEST SCORING, ITEM STATISTICS, AND ITEM FACTOR ANALYSIS

DISTRIBUTED BY
SCIENTIFIC SOFTWARE INTERNATONAL, INC.
7383 North Lincoln Avenue, Suite 100
Lincolnwood, IL 60712-1704
(847) 675-0720

PROGRAM COPYRIGHT HELD BY SCIENTIFIC SOFTWARE INTERNATIONAL, INC., 1985-2002
ALL RIGHTS RESERVED.

```
COMMAND FILE = DATA_1.TSF
OUTPUT FILE  = DATA_1.OUT
PHASE 0: INPUT COMMANDS

DATA_1.TSF - L1 TAKE L2 TEST DATA ADAPTIVE FULL-INFORMATION ITEM FACTOR
             ANALYSIS;    COUNTED RESPONSE PATTERNS
-----------------------------------------------------------------
>PROBLEM       NITEMS=30, RESPONSE=3;  <=== OMIT response should be included
>NAMES         ITEM1,ITEM2,ITEM3,ITEM4,ITEM5,ITEM6,ITEM7,ITEM8,ITEM9,ITEM10,
             ITEM11,ITEM12,ITEM13,ITEM14,ITEM15,ITEM16,ITEM17,ITEM18,ITEM19,
             ITEM20,ITEM21,ITEM22,ITEM23,ITEM24,ITEM25,ITEM26,ITEM27,ITEM28,
             ITEM29,ITEM30;
>RESPONSE      ' ','0','1';  <==== OMIT response comes first
>KEY           111111111111111111111111111111;
>RELIABILITY   ALPHA;
>TETRACHORIC   NDEC=3, LIST;
>FACTOR        NFAC=2, NIT=5, ROTATE=VARIMAX, RESIDUAL;
>FULL          CYCLES=100, CPARMS=(0.091,0.099,0.112,0.112,0.263,0.067,0.114,
             0.165,0.071,0.255,0.057,0.087,0.095,0.072,0.183,0.053,
             0.295,0.144,0.122,0.420,0.055,0.236,0.152,0.134,0.376,0.078,
             0.326,0.212,0.128,0.365), OMIT=RECODE;
>TECHNICAL     PRECISION=0.0001,QUAD=10;
>SCORE         METHOD=2;
```

```
>SAVE          CORR, ROTATE, PARM;
>INPUT         NIDCHAR=4, SCORES, FILE='DATA_1.DAT';

  DATA FILE NAME IS DATA_1.DAT

DATA FORMAT=
(4A1,1x,30A1)
```

PHASE 1 OF OUTPUT: DATA DESCRIPTION

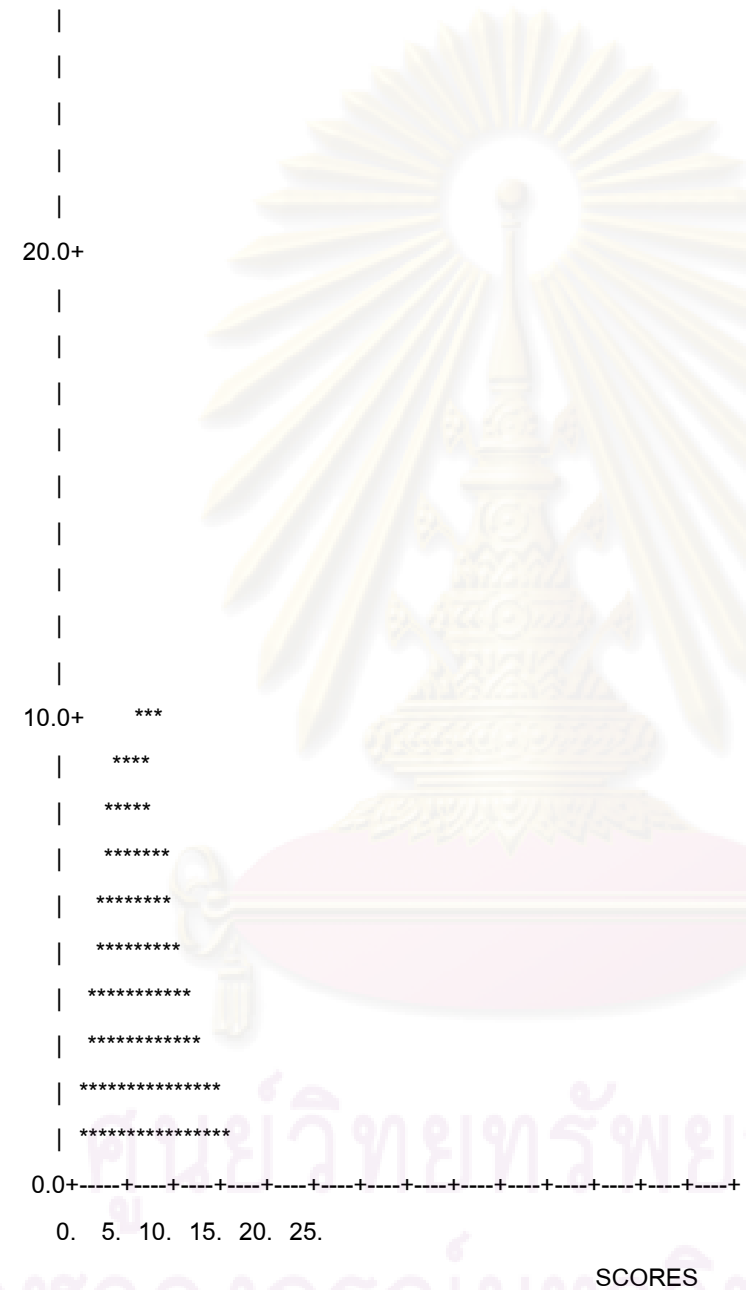
```
DATA_1.TSF - L1 TAKE L2 TEST DATA ADAPTIVE FULL-INFORMATION ITEM FACTOR
             ANALYSIS;    COUNTED RESPONSE PATTERNS
-------------------------------------------------------------
    RESPONSE CATEGORIES:   0 1
    ANSWER KEY: 11111111111111111111111111111

CONTENTS OF FIRST OBSERVATION:
    ID=   1
    WEIGHT=          1
    ITEM RESPONSES= 0100001000101001110101100011101

SUM OF WEIGHTS  =       2000
NUMBER OF RECORDS=      2000
```

PHASE 1: HISTOGRAM AND BASIC STATISTICS

```
DATA_1_EDIT.TSF - L1 TAKE L2 TEST DATA ADAPTIVE FULL-INFORMATION ITEM FACTOR
             ANALYSIS;    COUNTED RESPONSE PATTERNS
-------------------------------------------------------------
```

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

MAIN TEST HISTOGRAM



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

NUMBER OF OBSERVATIONS AT EACH SCORE

SCORE	COUNT	FREQ	SCORE	COUNT	FREQ	SCORE	COUNT	FREQ
0	1	0.1	11	148	7.4	22	3	0.2
1	16	0.8	12	109	5.4	23	3	0.2
2	42	2.1	13	89	4.4	24	0	0.0
3	87	4.3	14	76	3.8	25	0	0.0
4	134	6.7	15	50	2.5	26	0	0.0
5	173	8.6	16	48	2.4	27	0	0.0
6	200	10.0	17	25	1.2	28	0	0.0
7	201	10.1	18	15	0.8	29	0	0.0
8	212	10.6	19	12	0.6	30	0	0.0
9	195	9.8	20	8	0.4			
10	151	7.6	21	2	0.1			

TEST	RECORD	NUMBER	MEAN	S.D.	PROPORTION	S.D.
MAIN	2000	2000	8.6	3.9	0.288	0.453

PHASE 2: ITEM STATISTICS

DATA_1_EDIT.TSF - L1 TAKE L2 TEST DATA ADAPTIVE FULL-INFORMATION ITEM FACTOR ANALYSIS; COUNTED RESPONSE PATTERNS

MAIN TEST ITEM STATISTICS

ITEM	NUMBER	MEAN	S.D.	RMEAN	FACILITY	DIFF	BIS	P.BIS
1 ITEM1	2000	8.63	3.91	10.68	0.131	17.50	0.322	0.203
2 ITEM2	2000	8.63	3.91	10.48	0.394	14.07	0.484	0.381
3 ITEM3	2000	8.63	3.91	10.54	0.266	15.49	0.396	0.295
4 ITEM4	2000	8.63	3.91	10.52	0.184	16.60	0.335	0.230
5 ITEM5	2000	8.63	3.91	9.87	0.505	12.95	0.401	0.320
6 ITEM6	2000	8.63	3.91	11.21	0.101	18.10	0.377	0.221
7 ITEM7	2000	8.63	3.91	10.60	0.407	13.94	0.529	0.418
8 ITEM8	2000	8.63	3.91	10.50	0.299	15.11	0.411	0.311
9 ITEM9	2000	8.63	3.91	11.22	0.172	16.79	0.446	0.301

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

10 ITEM10	2000	8.63	3.91	9.79	0.497	13.03	0.370	0.295
11 ITEM11	2000	8.63	3.91	11.34	0.087	18.43	0.381	0.214
12 ITEM12	2000	8.63	3.91	10.88	0.380	14.22	0.575	0.451
13 ITEM13	2000	8.63	3.91	10.90	0.250	15.69	0.458	0.336
14 ITEM14	2000	8.63	3.91	11.38	0.172	16.79	0.474	0.320
15 ITEM15	2000	8.63	3.91	9.94	0.495	13.06	0.415	0.331
16 ITEM16	2000	8.63	3.91	11.29	0.069	18.93	0.353	0.185
17 ITEM17	2000	8.63	3.91	10.50	0.421	13.80	0.515	0.408
18 ITEM18	2000	8.63	3.91	11.00	0.257	15.61	0.483	0.356
19 ITEM19	2000	8.63	3.91	10.83	0.176	16.73	0.383	0.260
20 ITEM20	2000	8.63	3.91	9.74	0.512	12.87	0.364	0.290
21 ITEM21	2000	8.63	3.91	11.52	0.080	18.61	0.398	0.219
22 ITEM22	2000	8.63	3.91	10.66	0.389	14.12	0.527	0.414
23 ITEM23	2000	8.63	3.91	10.90	0.236	15.87	0.446	0.324
24 ITEM24	2000	8.63	3.91	10.66	0.182	16.64	0.356	0.244
25 ITEM25	2000	8.63	3.91	9.74	0.521	12.79	0.372	0.297
26 ITEM26	2000	8.63	3.91	10.63	0.094	18.27	0.286	0.164
27 ITEM27	2000	8.63	3.91	10.38	0.421	13.80	0.480	0.381
28 ITEM28	2000	8.63	3.91	10.56	0.275	15.39	0.407	0.305
29 ITEM29	2000	8.63	3.91	10.54	0.154	17.08	0.317	0.208
30 ITEM30	2000	8.63	3.91	9.79	0.503	12.96	0.373	0.297

MAIN TEST COEFFICIENT ALPHA = 0.742

EXAMINEES VARIANCE COMPONENT = 0.0113 DEGREES OF FREEDOM = 1999

RESIDUAL VARIANCE COMPONENT = 0.1712 DEGREES OF FREEDOM = 57971

PHASE 5: TETRACHORIC CORRELATIONS

DATA_1_EDIT.TSF - L1 TAKE L2 TEST DATA ADAPTIVE FULL-INFORMATION ITEM FACTOR ANALYSIS; COUNTED RESPONSE PATTERNS

MAIN TEST MISSING RESPONSE INFORMATION

ITEM	NUMBER OF CASES	PERCENT CORRECT	PERCENT OMITTED	PERCENT NOT REACHED	PERCENT NOT PRESENTED
1. ITEM1	2000	13.1	0.0	0.0	0.0
2. ITEM2	2000	39.5	0.0	0.0	0.0
3. ITEM3	2000	26.7	0.0	0.0	0.0
4. ITEM4	2000	18.4	0.0	0.0	0.0

5. ITEM5	2000	50.5	0.0	0.0	0.0
6. ITEM6	2000	10.1	0.0	0.0	0.0
7. ITEM7	2000	40.8	0.0	0.0	0.0
8. ITEM8	2000	29.8	0.0	0.0	0.0
9. ITEM9	2000	17.2	0.0	0.0	0.0
10. ITEM10	2000	49.7	0.0	0.0	0.0
11. ITEM11	2000	8.8	0.0	0.0	0.0
12. ITEM12	2000	38.0	0.0	0.0	0.0
13. ITEM13	2000	25.1	0.0	0.0	0.0
14. ITEM14	2000	17.2	0.0	0.0	0.0
15. ITEM15	2000	49.5	0.0	0.0	0.0
16. ITEM16	2000	6.9	0.0	0.0	0.0
17. ITEM17	2000	42.0	0.0	0.0	0.0
18. ITEM18	2000	25.7	0.0	0.0	0.0
19. ITEM19	2000	17.5	0.0	0.0	0.0
20. ITEM20	2000	51.2	0.0	0.0	0.0
21. ITEM21	2000	8.1	0.0	0.0	0.0
22. ITEM22	2000	39.0	0.0	0.0	0.0
23. ITEM23	2000	23.6	0.0	0.0	0.0
24. ITEM24	2000	18.1	0.0	0.0	0.0
25. ITEM25	2000	52.1	0.0	0.0	0.0
26. ITEM26	2000	9.4	0.0	0.0	0.0
27. ITEM27	2000	42.1	0.0	0.0	0.0
28. ITEM28	2000	27.6	0.0	0.0	0.0
29. ITEM29	2000	15.4	0.0	0.0	0.0
30. ITEM30	2000	50.3	0.0	0.0	0.0

-->ITEM PAIR(16, 10):
CELL FREQUENCIES ARE TOO SMALL FOR A MEANINGFUL RESULT;
RTET= 1.00 SUBSTITUTED.
-->ITEM PAIR(21, 16):
CELL FREQUENCIES ARE TOO SMALL FOR A MEANINGFUL RESULT;
RTET=-1.00 SUBSTITUTED.

.....

-->ITEM PAIR(30, 21):
CELL FREQUENCIES ARE TOO SMALL FOR A MEANINGFUL RESULT;
RTET= 1.00 SUBSTITUTED.

AVERAGE TETRACHORIC CORRELATION = 0.3283

STANDARD DEVIATION = 0.2420

NUMBER OF VALID ITEM PAIRS = 417

DISPLAY 1. TETRACHORIC CORRELATION MATRIX

		1 ITEM1	2 ITEM2	3 ITEM3	4 ITEM4	5 ITEM5	6 ITEM6
1	ITEM1	1.000					
2	ITEM2	0.274	1.000				
3	ITEM3	0.389	0.411	1.000			
4	ITEM4	0.384	0.394	0.357	1.000		
5	ITEM5	0.331	0.338	0.466	0.501	1.000	
6	ITEM6	0.429	0.274	0.347	0.312	0.365	1.000
7	ITEM7	0.596	0.487	0.376	0.323	0.358	0.561
8	ITEM8	0.305	0.565	0.301	0.253	0.362	0.326
9	ITEM9	0.471	0.367	0.322	0.368	0.324	0.381
10	ITEM10	0.255	0.358	0.335	0.369	0.124	0.397
11	ITEM11	0.517	0.406	0.488	0.271	0.247	0.697
12	ITEM12	0.468	0.470	0.395	0.384	0.371	0.444
13	ITEM13	0.338	0.432	0.297	0.416	0.333	0.384
14	ITEM14	0.312	0.362	0.315	0.369	0.344	0.274
15	ITEM15	0.617	0.377	0.312	0.353	0.383	0.290
16	ITEM16	0.122	0.406	0.191	0.381	0.406	0.231
17	ITEM17	0.297	0.232	0.303	0.119	0.236	0.598
18	ITEM18	0.359	0.092	0.023	0.028	0.171	0.308
19	ITEM19	-0.217	0.146	0.085	0.014	0.165	-0.030
20	ITEM20	0.121	0.092	0.236	-0.133	0.280	0.374
21	ITEM21	-0.131	0.208	0.133	-0.245	0.311	0.153
22	ITEM22	0.437	0.054	0.073	0.267	0.198	0.270
23	ITEM23	-0.030	0.200	-0.061	0.192	0.147	0.329
24	ITEM24	0.497	0.090	0.287	0.294	0.089	0.220
25	ITEM25	0.168	0.039	0.082	-0.141	0.066	0.090

26	ITEM26	0.365	-0.353	-1.000	0.374	-0.209	0.586
27	ITEM27	0.005	0.202	0.118	-0.239	0.244	0.458
28	ITEM28	-0.111	-0.018	-0.138	0.265	0.074	0.434
29	ITEM29	-1.000	-0.373	-1.000	-1.000	-1.000	0.549
30	ITEM30	-0.443	0.241	-0.004	-0.006	0.234	-0.392

		7	8	9	10	11	12
		ITEM7	ITEM8	ITEM9	ITEM10	ITEM11	ITEM12
7	ITEM7	1.000					
8	ITEM8	0.479	1.000				
9	ITEM9	0.391	0.313	1.000			
10	ITEM10	0.409	0.373	0.233	1.000		
11	ITEM11	0.509	0.413	0.476	0.131	1.000	
12	ITEM12	0.496	0.405	0.378	0.313	0.315	1.000
13	ITEM13	0.384	0.286	0.384	0.186	0.405	0.489
14	ITEM14	0.416	0.473	0.372	0.321	0.514	0.381
15	ITEM15	0.351	0.358	0.231	0.104	0.302	0.381
16	ITEM16	0.318	0.591	0.486	1.000	0.135	0.407
17	ITEM17	0.356	0.283	0.264	0.330	0.402	0.320
18	ITEM18	0.175	0.211	0.238	0.317	0.327	0.271
19	ITEM19	0.113	0.169	0.111	0.176	0.329	0.457
20	ITEM20	0.100	0.116	0.052	0.072	0.364	0.279
21	ITEM21	0.256	-0.046	0.268	0.091	0.385	0.306
22	ITEM22	0.258	0.196	0.282	0.092	0.172	0.319
23	ITEM23	0.268	0.074	0.279	0.024	0.213	0.316
24	ITEM24	0.141	0.192	0.232	0.427	0.083	0.333
25	ITEM25	0.215	0.204	0.126	0.049	0.343	0.230
26	ITEM26	0.270	0.283	0.156	-0.021	-0.107	1.000
27	ITEM27	0.148	0.266	0.381	0.348	0.213	0.284
28	ITEM28	0.170	0.087	0.341	0.282	0.360	0.273
29	ITEM29	0.302	0.113	0.194	-0.173	-1.000	-0.001
30	ITEM30	0.174	0.056	0.128	-0.035	0.372	0.258

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

		13	14	15	16	17	18
		ITEM13	ITEM14	ITEM15	ITEM16	ITEM17	ITEM18
13	ITEM13	1.000					
14	ITEM14	0.436	1.000				
15	ITEM15	0.370	0.318	1.000			
16	ITEM16	0.312	0.651	0.030	1.000		
17	ITEM17	0.238	0.324	0.138	0.562	1.000	
18	ITEM18	0.307	0.239	0.178	0.561	0.755	1.000
19	ITEM19	0.274	0.273	0.155	0.433	0.390	0.543
20	ITEM20	0.033	0.481	0.103	0.542	0.867	0.523
21	ITEM21	0.175	0.270	0.038	-1.000	0.907	0.659
22	ITEM22	0.142	0.174	0.238	0.638	0.857	0.688
23	ITEM23	0.035	0.307	0.139	0.736	0.680	0.535
24	ITEM24	0.162	-0.159	0.091	0.256	0.738	0.614
25	ITEM25	0.130	0.154	0.139	-0.061	0.785	0.563
26	ITEM26	0.177	-0.025	-0.334	-1.000	0.397	0.404
27	ITEM27	0.332	0.395	0.183	0.682	0.840	0.781
28	ITEM28	0.305	0.176	0.061	0.783	0.809	0.732
29	ITEM29	-0.043	0.047	-0.122	0.498	1.000	0.503
30	ITEM30	0.212	0.284	0.290	0.148	0.640	0.499

		19	20	21	22	23	24
		ITEM19	ITEM20	ITEM21	ITEM22	ITEM23	ITEM24
19	ITEM19	1.000					
20	ITEM20	0.557	1.000				
21	ITEM21	0.546	0.601	1.000			
22	ITEM22	0.820	0.664	0.650	1.000		
23	ITEM23	0.536	0.768	0.404	0.853	1.000	
24	ITEM24	0.134	0.477	0.518	0.685	0.672	1.000
25	ITEM25	0.560	0.594	0.756	0.651	0.396	0.475
26	ITEM26	0.617	0.755	0.396	1.000	0.661	0.252
27	ITEM27	0.715	0.790	0.820	0.932	0.704	0.770
28	ITEM28	0.700	0.515	0.726	0.877	0.766	0.507
29	ITEM29	0.663	1.000	0.583	1.000	0.746	0.605
30	ITEM30	0.607	0.524	1.000	0.500	0.641	0.290

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

		25	26	27	28	29	30
		ITEM25	ITEM26	ITEM27	ITEM28	ITEM29	ITEM30
25	ITEM25	1.000					
26	ITEM26	0.831	1.000				
27	ITEM27	0.725	1.000	1.000			
28	ITEM28	0.540	0.652	0.983	1.000		
29	ITEM29	1.000	0.763	1.000	0.756	1.000	
30	ITEM30	0.423	0.731	0.741	0.645	0.692	1.000

FORMAT-FREE SAVE OF THE TETRACHORIC CORRELATION MATRIX

PHASE 6: FACTOR ANALYSIS

DATA_1.TSF - L1 TAKE L2 TEST DATA ADAPTIVE FULL-INFORMATION ITEM FACTOR ANALYSIS; COUNTED RESPONSE PATTERNS

NUMBER OF FACTORS = 2

FULL INFORMATION

ADAPTIVE QUADRATURE

DISPLAY 2. INTER-ITEM CORRELATION MATRIX CORRECTED BY CENTROID METHOD

		1	2	3	4	5	6
		ITEM1	ITEM2	ITEM3	ITEM4	ITEM5	ITEM6
1	ITEM1	1.000					
2	ITEM2	0.274	1.000				
3	ITEM3	0.389	0.411	1.000			
4	ITEM4	0.384	0.394	0.357	1.000		

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

5	ITEM5	0.331	0.338	0.466	0.501	1.000	
6	ITEM6	0.429	0.274	0.347	0.312	0.365	1.000
7	ITEM7	0.596	0.487	0.376	0.323	0.358	0.561
8	ITEM8	0.305	0.565	0.301	0.253	0.362	0.326
9	ITEM9	0.471	0.367	0.322	0.368	0.324	0.381
10	ITEM10	0.255	0.358	0.335	0.369	0.124	0.397
11	ITEM11	0.517	0.406	0.488	0.271	0.247	0.697
12	ITEM12	0.468	0.470	0.395	0.384	0.371	0.444
13	ITEM13	0.338	0.432	0.297	0.416	0.333	0.384
14	ITEM14	0.312	0.362	0.315	0.369	0.344	0.274
15	ITEM15	0.617	0.377	0.312	0.353	0.383	0.290
16	ITEM16	0.122	0.406	0.191	0.381	0.406	0.231
17	ITEM17	0.297	0.232	0.303	0.119	0.236	0.598
18	ITEM18	0.359	0.092	0.023	0.028	0.171	0.308
19	ITEM19	-0.217	0.146	0.085	0.014	0.165	-0.030
20	ITEM20	0.121	0.092	0.236	-0.133	0.280	0.374
21	ITEM21	-0.131	0.208	0.133	-0.245	0.311	0.153
22	ITEM22	0.437	0.054	0.073	0.267	0.198	0.270
23	ITEM23	-0.030	0.200	-0.061	0.192	0.147	0.329
24	ITEM24	0.497	0.090	0.287	0.294	0.089	0.220
25	ITEM25	0.168	0.039	0.082	-0.141	0.066	0.090
26	ITEM26	0.365	-0.353	0.250	0.374	-0.209	0.586
27	ITEM27	0.005	0.202	0.118	-0.239	0.244	0.458
28	ITEM28	-0.111	-0.018	-0.138	0.265	0.074	0.434
29	ITEM29	0.277	-0.373	0.262	0.246	0.281	0.549
30	ITEM30	-0.443	0.241	-0.004	-0.006	0.234	-0.392

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

	7	8	9	10	11	12
	ITEM7	ITEM8	ITEM9	ITEM10	ITEM11	ITEM12
7	ITEM7	1.000				
8	ITEM8	0.479	1.000			
9	ITEM9	0.391	0.313	1.000		
10	ITEM10	0.409	0.373	0.233	1.000	

11	ITEM11	0.509	0.413	0.476	0.131	1.000	
12	ITEM12	0.496	0.405	0.378	0.313	0.315	1.000
13	ITEM13	0.384	0.286	0.384	0.186	0.405	0.489
14	ITEM14	0.416	0.473	0.372	0.321	0.514	0.381
15	ITEM15	0.351	0.358	0.231	0.104	0.302	0.381
16	ITEM16	0.318	0.591	0.486	0.276	0.135	0.407
17	ITEM17	0.356	0.283	0.264	0.330	0.402	0.320
18	ITEM18	0.175	0.211	0.238	0.317	0.327	0.271
19	ITEM19	0.113	0.169	0.111	0.176	0.329	0.457
20	ITEM20	0.100	0.116	0.052	0.072	0.364	0.279
21	ITEM21	0.256	-0.046	0.268	0.091	0.385	0.306
22	ITEM22	0.258	0.196	0.282	0.092	0.172	0.319
23	ITEM23	0.268	0.074	0.279	0.024	0.213	0.316
24	ITEM24	0.141	0.192	0.232	0.427	0.083	0.333
25	ITEM25	0.215	0.204	0.126	0.049	0.343	0.230
26	ITEM26	0.270	0.283	0.156	-0.021	-0.107	0.349
27	ITEM27	0.148	0.266	0.381	0.348	0.213	0.284
28	ITEM28	0.170	0.087	0.341	0.282	0.360	0.273
29	ITEM29	0.302	0.113	0.194	-0.173	0.350	-0.001
30	ITEM30	0.174	0.056	0.128	-0.035	0.372	0.258

	13	14	15	16	17	18
	ITEM13	ITEM14	ITEM15	ITEM16	ITEM17	ITEM18

13	ITEM13	1.000					
14	ITEM14	0.436	1.000				
15	ITEM15	0.370	0.318	1.000			
16	ITEM16	0.312	0.651	0.030	1.000		
17	ITEM17	0.238	0.324	0.138	0.562	1.000	
18	ITEM18	0.307	0.239	0.178	0.561	0.755	1.000
19	ITEM19	0.274	0.273	0.155	0.433	0.390	0.543
20	ITEM20	0.033	0.481	0.103	0.542	0.867	0.523
21	ITEM21	0.175	0.270	0.038	0.445	0.907	0.659
22	ITEM22	0.142	0.174	0.238	0.638	0.857	0.688

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

23	ITEM23	0.035	0.307	0.139	0.736	0.680	0.535
24	ITEM24	0.162	-0.159	0.091	0.256	0.738	0.614
25	ITEM25	0.130	0.154	0.139	-0.061	0.785	0.563
26	ITEM26	0.177	-0.025	-0.334	0.398	0.397	0.404
27	ITEM27	0.332	0.395	0.183	0.682	0.840	0.781
28	ITEM28	0.305	0.176	0.061	0.783	0.809	0.732
29	ITEM29	-0.043	0.047	-0.122	0.498	0.499	0.503
30	ITEM30	0.212	0.284	0.290	0.148	0.640	0.499

	19	20	21	22	23	24
	ITEM19	ITEM20	ITEM21	ITEM22	ITEM23	ITEM24
19	ITEM19	1.000				
20	ITEM20	0.557	1.000			
21	ITEM21	0.546	0.601	1.000		
22	ITEM22	0.820	0.664	0.650	1.000	
23	ITEM23	0.536	0.768	0.404	0.853	1.000
24	ITEM24	0.134	0.477	0.518	0.685	0.672	1.000
25	ITEM25	0.560	0.594	0.756	0.651	0.396	0.475
26	ITEM26	0.617	0.755	0.396	0.437	0.661	0.252
27	ITEM27	0.715	0.790	0.820	0.932	0.704	0.770
28	ITEM28	0.700	0.515	0.726	0.877	0.766	0.507
29	ITEM29	0.663	0.400	0.583	0.458	0.746	0.605
30	ITEM30	0.607	0.524	0.343	0.500	0.641	0.290

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

	25	26	27	28	29	30
	ITEM25	ITEM26	ITEM27	ITEM28	ITEM29	ITEM30
25	ITEM25	1.000				
26	ITEM26	0.831	1.000			

27	ITEM27	0.725	0.464	1.000			
28	ITEM28	0.540	0.652	0.983	1.000		
29	ITEM29	0.345	0.763	0.487	0.756	1.000	
30	ITEM30	0.423	0.731	0.741	0.645	0.692	1.000

DISPLAY 3. THE POSITIVE LATENT ROOTS OF THE CORRELATION MATRIX

	1	2	3	4	5	6
1	11.22053	4.93550	2.14518	1.82604	1.65743	1.38350

	7	8	9	10	11	12
1	1.28316	1.13950	1.06194	0.89442	0.81616	0.77777

	13	14	15	16	17	18
1	0.72798	0.63524	0.60190	0.54238	0.51283	0.37699

	19	20	21	22
1	0.27890	0.25939	0.16188	0.00814

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

DISPLAY 4. NUMBER OF ITEMS AND SUM OF LATENT ROOTS
AND THEIR RATIO
30 33.2467364 0.9023442

DISPLAY 5. THE CORRECTED LATENT ROOTS OF THE CORRELATION MATRIX

	1	2	3	4	5	6
1	10.12478	4.45352	1.93569	1.64771	1.49557	1.24839
	7	8	9	10	11	12
1	1.15785	1.02822	0.95823	0.80708	0.73646	0.70182
	13	14	15	16	17	18
1	0.65688	0.57321	0.54312	0.48942	0.46275	0.34017
	19	20	21	22		
1	0.25167	0.23406	0.14607	0.00734		

DISPLAY 6. INITIAL SMOOTHED INTER-ITEM CORRELATION MATRIX

		1	2	3	4	5	6
		ITEM1	ITEM2	ITEM3	ITEM4	ITEM5	ITEM6
1	ITEM1	1.000					
2	ITEM2	0.171	1.000				
3	ITEM3	0.405	0.372	1.000			
4	ITEM4	0.395	0.307	0.351	1.000		
5	ITEM5	0.257	0.353	0.433	0.406	1.000	
6	ITEM6	0.449	0.179	0.344	0.295	0.282	1.000
7	ITEM7	0.528	0.464	0.365	0.323	0.340	0.499
8	ITEM8	0.331	0.510	0.318	0.270	0.318	0.310
9	ITEM9	0.414	0.366	0.313	0.342	0.310	0.344
10	ITEM10	0.206	0.358	0.310	0.318	0.120	0.328
11	ITEM11	0.377	0.403	0.417	0.209	0.275	0.530
12	ITEM12	0.396	0.470	0.375	0.364	0.347	0.360
13	ITEM13	0.303	0.425	0.289	0.386	0.312	0.335
14	ITEM14	0.235	0.381	0.293	0.282	0.364	0.227
15	ITEM15	0.475	0.397	0.266	0.282	0.399	0.201
16	ITEM16	0.113	0.330	0.160	0.332	0.352	0.282
17	ITEM17	0.230	0.241	0.257	0.092	0.253	0.439
18	ITEM18	0.304	0.100	0.021	0.032	0.165	0.260
19	ITEM19	-0.120	0.090	0.097	0.038	0.148	0.042
20	ITEM20	0.109	0.084	0.239	-0.078	0.214	0.337
21	ITEM21	-0.033	0.137	0.152	-0.177	0.259	0.196
22	ITEM22	0.293	0.098	0.044	0.181	0.194	0.209
23	ITEM23	0.057	0.153	-0.029	0.204	0.130	0.282

24	ITEM24	0.381	0.099	0.248	0.199	0.125	0.228
25	ITEM25	0.175	0.015	0.087	-0.111	0.020	0.147
26	ITEM26	0.160	-0.220	0.157	0.207	-0.075	0.340
27	ITEM27	0.009	0.160	0.095	-0.124	0.203	0.300
28	ITEM28	-0.064	0.004	-0.088	0.201	0.087	0.367
29	ITEM29	0.143	-0.264	0.207	0.196	0.238	0.362
30	ITEM30	-0.247	0.127	0.037	0.038	0.148	-0.152

		7	8	9	10	11	12
		ITEM7	ITEM8	ITEM9	ITEM10	ITEM11	ITEM12
7	ITEM7	1.000					
8	ITEM8	0.457	1.000				
9	ITEM9	0.390	0.303	1.000			
10	ITEM10	0.365	0.347	0.223	1.000		
11	ITEM11	0.475	0.333	0.438	0.143	1.000	
12	ITEM12	0.481	0.383	0.376	0.322	0.303	1.000
13	ITEM13	0.361	0.282	0.381	0.205	0.375	0.493
14	ITEM14	0.401	0.433	0.370	0.297	0.489	0.356
15	ITEM15	0.338	0.299	0.227	0.116	0.340	0.377
16	ITEM16	0.283	0.482	0.435	0.252	0.135	0.357
17	ITEM17	0.340	0.234	0.250	0.302	0.410	0.308
18	ITEM18	0.186	0.209	0.242	0.299	0.309	0.271
19	ITEM19	0.123	0.166	0.103	0.107	0.261	0.386
20	ITEM20	0.111	0.140	0.068	0.064	0.274	0.288
21	ITEM21	0.222	0.010	0.240	0.084	0.313	0.261
22	ITEM22	0.216	0.162	0.273	0.137	0.184	0.317
23	ITEM23	0.233	0.113	0.248	0.028	0.176	0.289
24	ITEM24	0.147	0.132	0.223	0.354	0.098	0.271
25	ITEM25	0.196	0.176	0.120	0.055	0.267	0.227
26	ITEM26	0.227	0.176	0.131	0.002	0.046	0.296
27	ITEM27	0.157	0.244	0.333	0.304	0.229	0.281
28	ITEM28	0.165	0.112	0.324	0.227	0.265	0.238
29	ITEM29	0.262	0.083	0.195	-0.096	0.318	0.066
30	ITEM30	0.136	0.083	0.104	-0.047	0.231	0.189

		13	14	15	16	17	18
		ITEM13	ITEM14	ITEM15	ITEM16	ITEM17	ITEM18
13	ITEM13	1.000					
14	ITEM14	0.412	1.000				
15	ITEM15	0.357	0.328	1.000			
16	ITEM16	0.275	0.560	0.031	1.000		
17	ITEM17	0.216	0.311	0.163	0.471	1.000	
18	ITEM18	0.301	0.239	0.178	0.487	0.709	1.000
19	ITEM19	0.225	0.251	0.111	0.408	0.368	0.500
20	ITEM20	0.049	0.430	0.069	0.447	0.749	0.517
21	ITEM21	0.166	0.221	-0.011	0.382	0.767	0.612
22	ITEM22	0.159	0.187	0.257	0.546	0.727	0.651
23	ITEM23	0.039	0.256	0.094	0.613	0.622	0.494
24	ITEM24	0.117	-0.086	0.113	0.265	0.658	0.572
25	ITEM25	0.134	0.110	0.104	0.028	0.666	0.515
26	ITEM26	0.137	0.051	-0.183	0.287	0.409	0.358
27	ITEM27	0.295	0.301	0.167	0.548	0.788	0.719
28	ITEM28	0.276	0.200	0.030	0.599	0.706	0.688
29	ITEM29	0.006	0.038	-0.070	0.379	0.452	0.470
30	ITEM30	0.182	0.210	0.179	0.210	0.452	0.418

		19	20	21	22	23	24
		ITEM19	ITEM20	ITEM21	ITEM22	ITEM23	ITEM24
19	ITEM19	1.000					
20	ITEM20	0.482	1.000				
21	ITEM21	0.511	0.547	1.000			
22	ITEM22	0.643	0.608	0.570	1.000		
23	ITEM23	0.500	0.691	0.424	0.723	1.000	
24	ITEM24	0.164	0.383	0.426	0.609	0.538	1.000
25	ITEM25	0.506	0.572	0.667	0.575	0.338	0.409
26	ITEM26	0.466	0.534	0.266	0.417	0.514	0.282
27	ITEM27	0.614	0.695	0.715	0.747	0.655	0.584
28	ITEM28	0.641	0.480	0.633	0.742	0.704	0.459
29	ITEM29	0.467	0.405	0.437	0.468	0.614	0.446
30	ITEM30	0.581	0.462	0.381	0.391	0.573	0.249

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

		25	26	27	28	29	30
		ITEM25	ITEM26	ITEM27	ITEM28	ITEM29	ITEM30
25	ITEM25	1.000					
26	ITEM26	0.577	1.000				
27	ITEM27	0.596	0.380	1.000			
28	ITEM28	0.461	0.509	0.785	1.000		
29	ITEM29	0.319	0.615	0.428	0.618	1.000	
30	ITEM30	0.438	0.412	0.539	0.530	0.465	1.000

DISPLAY 7. ITERATED COMMUNALITY ESTIMATES

		1	2	3	4	5
1	ITEM1	0.467	0.390	0.379	0.377	0.377
2	ITEM2	0.452	0.433	0.432	0.433	0.433
3	ITEM3	0.356	0.341	0.341	0.341	0.341
4	ITEM4	0.367	0.318	0.314	0.313	0.313
5	ITEM5	0.329	0.306	0.305	0.305	0.305
6	ITEM6	0.419	0.347	0.341	0.340	0.340
7	ITEM7	0.482	0.485	0.488	0.489	0.489
8	ITEM8	0.388	0.369	0.368	0.369	0.369
9	ITEM9	0.357	0.353	0.353	0.354	0.354
10	ITEM10	0.236	0.211	0.210	0.210	0.210
11	ITEM11	0.415	0.380	0.377	0.377	0.377
12	ITEM12	0.440	0.445	0.447	0.448	0.448
13	ITEM13	0.353	0.347	0.349	0.349	0.349
14	ITEM14	0.380	0.365	0.365	0.365	0.365
15	ITEM15	0.319	0.289	0.286	0.286	0.286
16	ITEM16	0.471	0.420	0.417	0.417	0.417
17	ITEM17	0.757	0.756	0.759	0.760	0.760
18	ITEM18	0.556	0.582	0.589	0.590	0.590
19	ITEM19	0.509	0.466	0.462	0.462	0.462
20	ITEM20	0.589	0.569	0.569	0.569	0.569

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

21	ITEM21	0.574	0.549	0.549	0.549	0.549
22	ITEM22	0.718	0.693	0.693	0.693	0.693
23	ITEM23	0.639	0.618	0.617	0.617	0.617
24	ITEM24	0.398	0.355	0.354	0.354	0.354
25	ITEM25	0.491	0.451	0.448	0.448	0.448
26	ITEM26	0.497	0.365	0.353	0.352	0.352
27	ITEM27	0.841	0.797	0.793	0.792	0.792
28	ITEM28	0.785	0.752	0.749	0.748	0.748
29	ITEM29	0.514	0.425	0.417	0.416	0.416
30	ITEM30	0.478	0.378	0.368	0.367	0.367

DISPLAY 8. THE NROOT LARGEST LATENT ROOTS OF THE CORRELATION MATRIX

	1	2
1	9.553320	3.935452

DISPLAY 10. INITIAL CHANCE, INTERCEPT, AND SLOPE ESTIMATES

		CHANCE	INTERCEPT	SLOPES	
			1	2	
1	ITEM1	0.091	-2.165	0.018	0.777
2	ITEM2	0.099	-0.591	-0.030	0.873
3	ITEM3	0.112	-1.156	0.037	0.718
4	ITEM4	0.112	-1.686	-0.015	0.675
5	ITEM5	0.263	-0.533	0.134	0.649
6	ITEM6	0.067	-2.203	0.297	0.653
7	ITEM7	0.114	-0.610	0.197	0.958
8	ITEM8	0.165	-1.252	0.129	0.753
9	ITEM9	0.071	-1.534	0.239	0.700
10	ITEM10	0.255	-0.511	0.116	0.502
11	ITEM11	0.057	-2.335	0.282	0.725
12	ITEM12	0.087	-0.626	0.327	0.839
13	ITEM13	0.095	-1.174	0.157	0.715
14	ITEM14	0.072	-1.555	0.240	0.720

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

15	ITEM15	0.183	-0.358	0.014	0.633
16	ITEM16	0.053	-2.767	0.655	0.536
17	ITEM17	0.295	-1.885	1.666	0.630
18	ITEM18	0.144	-1.744	1.150	0.343
19	ITEM19	0.122	-2.108	0.919	0.121
20	ITEM20	0.420	-1.518	1.130	0.204
21	ITEM21	0.055	-2.860	1.093	0.147
22	ITEM22	0.236	-1.513	1.463	0.343
23	ITEM23	0.152	-2.074	1.253	0.202
24	ITEM24	0.134	-1.989	0.697	0.248
25	ITEM25	0.376	-0.984	0.898	0.073
26	ITEM26	0.078	-2.610	0.735	0.051
27	ITEM27	0.326	-2.359	1.914	0.378
28	ITEM28	0.212	-2.792	1.712	0.204
29	ITEM29	0.128	-2.458	0.840	0.083
30	ITEM30	0.365	-0.979	0.762	0.018

DISPLAY 11. THE EM ESTIMATION OF PARAMETERS

10 QUADRATURE POINTS

DISPLAY 12. 10 QUADRATURE POINTS AND WEIGHTS:

1	-4.859463	0.000004
2	-3.581823	0.000758
3	-2.484326	0.019112
4	-1.465989	0.135484
5	-0.484936	0.344642
6	0.484936	0.344642
7	1.465989	0.135484
8	2.484326	0.019112
9	3.581823	0.000758
10	4.859463	0.000004

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

CYCLE 1 - 2 X MARGINAL LOG LIKELIHOOD = 0.6279354757D+05

MAXIMUM CHANGE OF ESTIMATES

INTERCEPT = 0.691405 SLOPE = 0.446224

0.115851

CYCLE 99 - 2 X MARGINAL LOG LIKELIHOOD = 0.6272404511D+05

CHANGE = 0.1915188222D-03

MAXIMUM CHANGE OF ESTIMATES

INTERCEPT = 0.000029 SLOPE = 0.000162

0.000303

DISPLAY 14. UNTRANSFORMED ITEM PARAMETERS

CHANCE INTERCEPT SLOPE ESTIMATES

		1	2	
1	ITEM1	0.091	-2.224	0.156	0.683
2	ITEM2	0.099	-0.721	0.017	0.869
3	ITEM3	0.112	-1.252	-0.024	0.704
4	ITEM4	0.112	-1.764	-0.020	0.661
5	ITEM5	0.263	-0.642	0.115	0.606
6	ITEM6	0.067	-2.629	0.340	0.807
7	ITEM7	0.114	-0.756	0.167	0.859
8	ITEM8	0.165	-1.457	0.159	0.793
9	ITEM9	0.071	-1.613	0.173	0.617
10	ITEM10	0.255	-0.599	0.138	0.439
11	ITEM11	0.057	-2.815	0.253	0.937
12	ITEM12	0.087	-0.810	0.269	0.807
13	ITEM13	0.095	-1.308	0.140	0.691
14	ITEM14	0.072	-1.695	0.233	0.670
15	ITEM15	0.183	-0.445	0.078	0.561
16	ITEM16	0.053	-3.398	0.896	0.415
17	ITEM17	0.295	-1.980	1.280	0.443
18	ITEM18	0.144	-1.858	0.917	0.197
19	ITEM19	0.122	-2.165	0.727	0.177
20	ITEM20	0.420	-1.642	0.869	0.222

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

21	ITEM21	0.055	-2.626	0.752	0.158
22	ITEM22	0.236	-1.799	1.299	0.178
23	ITEM23	0.152	-2.097	0.933	0.225
24	ITEM24	0.134	-2.178	0.680	0.189
25	ITEM25	0.376	-1.154	0.745	0.057
26	ITEM26	0.078	-3.418	1.073	-0.040
27	ITEM27	0.326	-2.508	1.588	0.353
28	ITEM28	0.212	-2.276	1.026	0.159
29	ITEM29	0.128	-2.618	0.913	-0.101
30	ITEM30	0.365	-1.138	0.621	0.175

DISPLAY 17. RESIDUAL CORRELATIONS

		1	2	3	4	5	6
		ITEM1	ITEM2	ITEM3	ITEM4	ITEM5	ITEM6
1	ITEM1	0.671					
2	ITEM2	-0.198	0.570				
3	ITEM3	0.086	-0.005	0.668			
4	ITEM4	0.089	-0.054	0.033	0.695		
5	ITEM5	-0.044	0.013	0.138	0.123	0.724	
6	ITEM6	0.077	-0.222	-0.001	-0.036	-0.056	0.566
7	ITEM7	0.150	0.039	-0.005	-0.032	-0.006	0.074
8	ITEM8	-0.029	0.104	-0.035	-0.068	-0.012	-0.096
9	ITEM9	0.105	0.023	0.017	0.058	0.028	-0.008
10	ITEM10	-0.033	0.095	0.083	0.100	-0.098	0.054
11	ITEM11	-0.022	-0.040	0.034	-0.159	-0.089	0.075
12	ITEM12	0.025	0.064	0.024	0.028	0.010	-0.066
13	ITEM13	-0.028	0.053	-0.034	0.077	0.009	-0.037
14	ITEM14	-0.095	0.020	-0.018	-0.016	0.064	-0.153
15	ITEM15	0.194	0.075	-0.014	0.014	0.141	-0.113
16	ITEM16	-0.134	0.128	0.002	0.180	0.138	-0.061
17	ITEM17	-0.015	0.059	0.120	-0.040	0.043	0.085
18	ITEM18	0.138	-0.003	-0.049	-0.036	0.025	0.002
19	ITEM19	-0.274	-0.010	0.027	-0.030	0.018	-0.193
20	ITEM20	-0.067	-0.033	0.156	-0.159	0.066	0.071
21	ITEM21	-0.180	0.048	0.091	-0.236	0.136	-0.033
22	ITEM22	0.132	0.017	-0.003	0.135	0.062	-0.058

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

23	ITEM23	-0.120	0.038	-0.110	0.126	-0.019	0.012
24	ITEM24	0.224	-0.010	0.170	0.123	-0.009	-0.008
25	ITEM25	0.073	-0.023	0.072	-0.126	-0.062	-0.033
26	ITEM26	0.081	-0.211	0.187	0.234	-0.133	0.169
27	ITEM27	-0.201	0.028	0.004	-0.212	0.026	-0.025
28	ITEM28	-0.217	-0.077	-0.138	0.153	-0.040	0.118
29	ITEM29	0.098	-0.224	0.263	0.249	0.211	0.235
30	ITEM30	-0.396	0.025	-0.037	-0.035	0.021	-0.374

		7	8	9	10	11	12
		ITEM7	ITEM8	ITEM9	ITEM10	ITEM11	ITEM12
7	ITEM7	0.566					
8	ITEM8	0.043	0.605				
9	ITEM9	0.035	-0.035	0.709			
10	ITEM10	0.091	0.086	-0.002	0.826		
11	ITEM11	0.017	-0.103	0.062	-0.148	0.515	
12	ITEM12	0.058	-0.021	0.027	0.051	-0.147	0.580
13	ITEM13	-0.018	-0.080	0.071	-0.034	-0.026	0.122
14	ITEM14	0.024	0.073	0.059	0.056	0.087	-0.019
15	ITEM15	0.013	-0.011	-0.036	-0.087	-0.001	0.063
16	ITEM16	0.012	0.221	0.189	0.055	-0.180	0.045
17	ITEM17	0.074	-0.022	0.003	0.102	0.095	-0.010
18	ITEM18	0.009	0.037	0.070	0.158	0.091	0.046
19	ITEM19	-0.041	0.007	-0.056	-0.022	0.060	0.180
20	ITEM20	-0.077	-0.042	-0.112	-0.083	0.045	0.054
21	ITEM21	0.066	-0.141	0.089	-0.041	0.120	0.062
22	ITEM22	0.047	-0.002	0.102	-0.005	-0.032	0.089
23	ITEM23	0.043	-0.070	0.066	-0.121	-0.055	0.051
24	ITEM24	-0.022	-0.032	0.062	0.224	-0.106	0.063
25	ITEM25	0.091	0.074	0.009	-0.038	0.128	0.076
26	ITEM26	0.153	0.102	0.038	-0.078	-0.069	0.163
27	ITEM27	-0.068	0.027	0.116	0.127	-0.046	-0.003
28	ITEM28	0.004	-0.044	0.163	0.094	0.062	0.025
29	ITEM29	0.226	0.046	0.136	-0.150	0.246	-0.026
30	ITEM30	-0.024	-0.072	-0.048	-0.171	0.038	-0.009

		13	14	15	16	17	18
		ITEM13	ITEM14	ITEM15	ITEM16	ITEM17	ITEM18
13	ITEM13	0.668					
14	ITEM14	0.082	0.665				
15	ITEM15	0.074	0.048	0.757			
16	ITEM16	0.035	0.278	-0.156	0.507		
17	ITEM17	-0.019	0.023	-0.017	-0.091	0.353	
18	ITEM18	0.143	0.034	0.063	0.018	0.162	0.532
19	ITEM19	0.078	0.064	0.002	-0.005	-0.112	0.090
20	ITEM20	-0.119	0.217	-0.056	-0.014	0.213	0.060
21	ITEM21	0.027	0.039	-0.113	-0.036	0.280	0.195
22	ITEM22	0.008	-0.021	0.151	0.012	0.100	0.108
23	ITEM23	-0.130	0.040	-0.031	0.136	0.068	0.021
24	ITEM24	-0.034	-0.275	0.000	-0.134	0.195	0.178
25	ITEM25	0.039	-0.029	0.041	-0.366	0.201	0.110
26	ITEM26	0.069	-0.073	-0.219	-0.171	-0.140	-0.128
27	ITEM27	0.095	0.042	0.021	-0.036	0.107	0.136
28	ITEM28	0.132	0.005	-0.072	0.113	0.136	0.196
29	ITEM29	-0.029	-0.049	-0.080	-0.027	-0.040	0.031
30	ITEM30	0.039	0.031	0.072	-0.166	0.017	0.047

		19	20	21	22	23	24
		ITEM19	ITEM20	ITEM21	ITEM22	ITEM23	ITEM24
19	ITEM19	0.641					
20	ITEM20	0.082	0.554				
21	ITEM21	0.146	0.140	0.629			
22	ITEM22	0.169	0.080	0.087	0.368		
23	ITEM23	0.085	0.228	0.003	0.175	0.521	
24	ITEM24	-0.181	-0.002	0.076	0.155	0.139	0.668
25	ITEM25	0.152	0.179	0.305	0.100	-0.071	0.070
26	ITEM26	0.044	0.066	-0.167	-0.157	0.026	-0.120
27	ITEM27	0.103	0.127	0.196	0.071	0.065	0.093
28	ITEM28	0.211	0.001	0.194	0.169	0.208	0.047
29	ITEM29	0.086	-0.017	0.045	-0.053	0.174	0.084
30	ITEM30	0.256	0.099	0.051	-0.036	0.198	-0.063

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

		25	26	27	28	29	30
		ITEM25	ITEM26	ITEM27	ITEM28	ITEM29	ITEM30
25	ITEM25	0.642					
26	ITEM26	0.142	0.464				
27	ITEM27	0.092	-0.223	0.274			
28	ITEM28	0.032	-0.009	0.173	0.481		
29	ITEM29	-0.078	0.121	-0.118	0.147	0.542	
30	ITEM30	0.120	0.035	0.078	0.142	0.125	0.706

DISPLAY 19. 9 FACTOR SCORE QUADRATURE POINTS AND WEIGHTS:

1	-4.512746	0.000022
2	-3.205429	0.002789
3	-2.076848	0.049916
4	-1.023256	0.244098
5	0.000000	0.406349
6	1.023256	0.244098
7	2.076848	0.049916
8	3.205429	0.002789
9	4.512746	0.000022

1 30 43.3 0.0 1
1.854 0.355
0.541* 0.618*

:

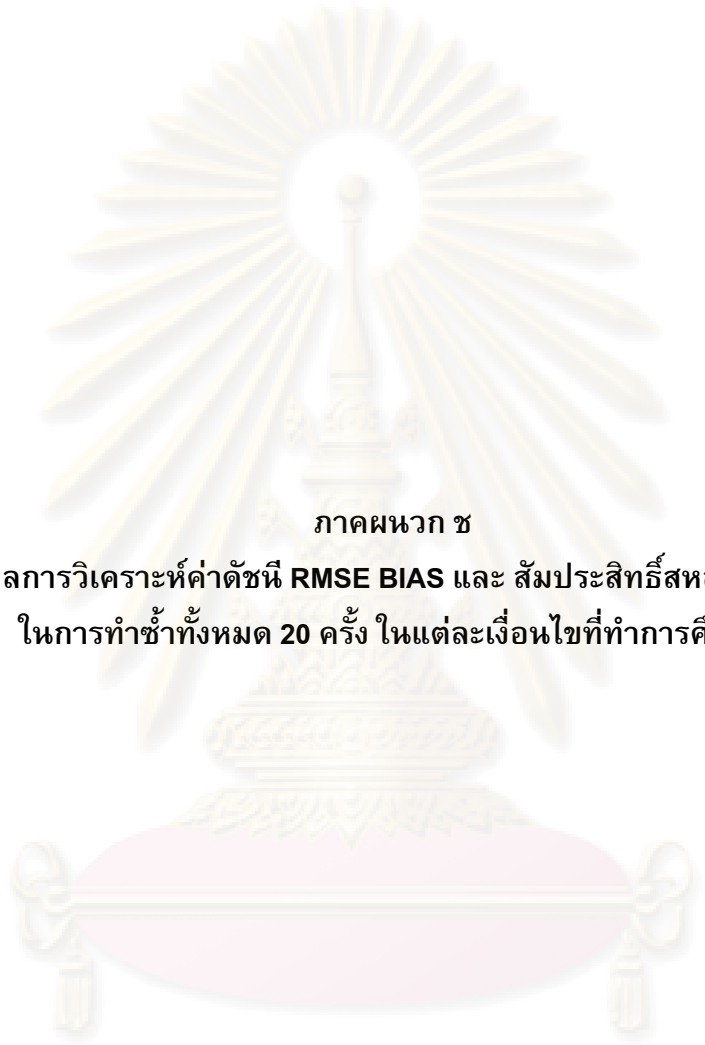
2000 30 20.0 0.0 2000
0.145 -0.813
0.990* 0.715*

ROOT-MEAN-SQUARE POSTERIOR STANDARD DEVIATIONS

RMS:	0.7970	0.6780
VARIANCE:	0.6352	0.4597

FORMATTED SAVE OF THE ROTATED FACTOR LOADINGS
FORMAT IS (I5,1X,A8,1X,5F10.5,/15X,5F10.5,/15X,5F10.5)
N O R M A L E N D O F T H I S P R O B L E M

START DATE: 2- 1-2008
START TIME: 10:32:08
END TIME: 10:41:34
NORMAL END



ภาคผนวก ช

ผลการวิเคราะห์ค่าดัชนี RMSE BIAS และ สัมประสิทธิ์สหสัมพันธ์

ในการทำซ้ำทั้งหมด 20 ครั้ง ในแต่ละเงื่อนไขที่ทำการศึกษา

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางภาคผนวกที่ 1 ค่าดัชนี RMSE BIAS และสัมประสิทธิ์สหสัมพันธ์ของเงื่อนไขไม่มีความสัมพันธ์ระหว่างมิติความสามารถในโครงสร้างที่ไม่ซับซ้อนของวิธี M และ วิธี NOP สำหรับการแปลงคะแนนจากระดับชั้นที่ 1 ไป 2 จากการทำซ้ำ 20 ครั้ง

ลำดับในการทำซ้ำ	L_12 CORR 0.0 APSS																	
	APSS_M									APSS_NOP								
	RMSE			BIAS_TRUE			Correlation Coefficient			RMSE			BIAS_TRUE			Correlation Coefficient		
	a1	a2	d	a1	a2	d	a1	a2	d	a1	a1	d	a1	a1	d	a1	a1	d
1	0.142	0.133	0.205	-0.074	-0.141	0.149	0.930	0.923	0.956	0.125	0.132	0.205	-0.107	-0.135	0.149	0.945	0.920	0.956
2	0.094	0.116	0.282	-0.128	-0.116	0.235	0.960	0.949	0.940	0.093	0.115	0.282	-0.123	-0.107	0.235	0.960	0.947	0.940
3	0.17	0.129	0.328	-0.113	-0.149	0.187	0.895	0.927	0.904	0.17	0.128	0.328	-0.112	-0.14	0.187	0.895	0.926	0.904
4	0.177	0.123	0.259	-0.102	-0.116	0.181	0.885	0.938	0.941	0.164	0.122	0.259	-0.135	-0.123	0.181	0.903	0.939	0.941
5	0.126	0.183	0.27	-0.139	-0.098	0.154	0.919	0.890	0.928	0.125	0.182	0.27	-0.133	-0.087	0.154	0.918	0.888	0.928
6	0.188	0.134	0.274	-0.059	-0.145	0.142	0.896	0.925	0.934	0.169	0.133	0.274	-0.098	-0.134	0.142	0.915	0.922	0.934
7	0.144	0.133	0.274	-0.103	-0.141	0.198	0.917	0.916	0.934	0.143	0.132	0.274	-0.114	-0.136	0.198	0.921	0.914	0.934
8	0.202	0.111	0.269	-0.062	-0.122	0.145	0.874	0.949	0.930	0.178	0.109	0.269	-0.106	-0.133	0.145	0.901	0.951	0.930
9	0.231	0.117	0.306	-0.069	-0.139	0.146	0.850	0.933	0.933	0.221	0.117	0.306	-0.105	-0.142	0.146	0.866	0.934	0.933
10	0.16	0.139	0.222	-0.086	-0.142	0.165	0.899	0.920	0.925	0.147	0.136	0.222	-0.114	-0.126	0.165	0.923	0.930	0.925
11	0.118	0.134	0.263	-0.13	-0.079	0.205	0.923	0.930	0.925	0.118	0.134	0.263	-0.136	-0.077	0.205	0.925	0.930	0.925
12	0.189	0.118	0.327	-0.135	-0.117	0.188	0.863	0.941	0.896	0.188	0.118	0.327	-0.141	-0.122	0.188	0.866	0.942	0.896
13	0.168	0.156	0.257	-0.111	-0.135	0.158	0.888	0.910	0.940	0.163	0.156	0.257	-0.132	-0.13	0.158	0.896	0.908	0.940
14	0.138	0.153	0.277	-0.127	-0.136	0.24	0.921	0.906	0.929	0.136	0.152	0.277	-0.14	-0.128	0.24	0.929	0.904	0.929
15	0.113	0.154	0.337	-0.106	-0.107	0.023	0.948	0.911	0.930	0.179	0.153	0.337	-0.045	-0.121	0.023	0.954	0.918	0.930
16	0.14	0.18	0.263	-0.149	-0.069	0.101	0.915	0.905	0.944	0.139	0.177	0.263	-0.14	-0.086	0.101	0.913	0.909	0.944
17	0.186	0.156	0.337	-0.018	-0.137	-0.006	0.909	0.870	0.917	0.179	0.153	0.337	-0.045	-0.121	-0.006	0.918	0.865	0.917
18	0.131	0.176	0.23	-0.133	-0.084	0.135	0.916	0.904	0.936	0.13	0.176	0.23	-0.142	-0.075	0.135	0.920	0.903	0.936
19	0.153	0.155	0.212	-0.1	-0.117	0.155	0.918	0.900	0.955	0.149	0.155	0.212	-0.12	-0.118	0.155	0.925	0.900	0.955
20	0.165	0.151	0.232	-0.109	-0.122	0.213	0.890	0.898	0.950	0.164	0.151	0.232	-0.117	-0.121	0.213	0.890	0.898	0.950
MEAN	0.157	0.143	0.271	-0.103	-0.121	0.156	0.906	0.917	0.932	0.154	0.142	0.271	-0.115	-0.118	0.156	0.914	0.918	0.932
SD	0.033	0.022	0.040	0.033	0.023	0.061	0.027	0.020	0.015	0.029	0.021	0.040	0.028	0.021	0.061	0.025	0.021	0.015
Min	0.094	0.111	0.205	-0.149	-0.149	-0.006	0.850	0.870	0.896	0.093	0.109	0.205	-0.142	-0.142	-0.006	0.866	0.865	0.896
Max	0.231	0.183	0.337	-0.018	-0.069	0.240	0.960	0.949	0.956	0.221	0.182	0.337	-0.045	-0.075	0.240	0.960	0.951	0.956

ตารางภาคผนวกที่ 2 ค่าดัชนี RMSE BIAS และสัมประสิทธิ์สหสัมพันธ์ของเงื่อนไขไม่มีความสัมพันธ์ระหว่างมิติความสามารถในโครงสร้างที่ซับซ้อนของวิธี M และ วิธี NOP สำหรับการแปลงคะแนนจากระดับชั้นที่ 1 ไป 2 จากการทำซ้ำ 20 ครั้ง

ลำดับในการทำซ้ำ	L_12 CORR 0.0 MS																	
	MS_M									MS_NOP								
	RMSE			BIAS_TRUE			Correlation Coefficient			RMSE			BIAS_TRUE			Correlation Coefficient		
	a1	a2	d	a1	a2	d	a1	a2	d	a1	a1	d	a1	a1	d	a1	a1	d
1	0.262	0.197	0.557	-0.178	-0.174	0.07	0.716	0.797	0.682	0.243	0.190	0.557	-0.218	-0.196	0.070	0.787	0.826	0.682
2	0.127	0.113	0.215	-0.157	-0.133	0.319	0.895	0.909	0.966	0.127	0.104	0.215	-0.158	-0.149	0.319	0.895	0.921	0.966
3	0.14	0.128	0.301	-0.143	-0.147	0.342	0.850	0.886	0.918	0.140	0.127	0.301	-0.144	-0.141	0.342	0.851	0.882	0.918
4	0.131	0.165	0.194	-0.157	-0.182	0.36	0.889	0.802	0.961	0.128	0.157	0.194	-0.166	-0.157	0.360	0.888	0.793	0.961
5	0.13	0.193	0.25	-0.181	-0.139	0.319	0.900	0.723	0.939	0.128	0.193	0.250	-0.172	-0.145	0.319	0.896	0.734	0.939
6	0.152	0.163	0.259	-0.161	-0.17	0.339	0.861	0.822	0.941	0.152	0.161	0.259	-0.163	-0.157	0.339	0.862	0.821	0.941
7	0.227	0.186	0.262	-0.171	-0.156	0.198	0.777	0.820	0.932	0.214	0.180	0.262	-0.203	-0.174	0.198	0.803	0.807	0.932
8	0.121	0.134	0.125	-0.179	-0.135	0.344	0.914	0.859	0.982	0.120	0.134	0.125	-0.186	-0.137	0.344	0.913	0.853	0.982
9	0.126	0.148	0.22	-0.205	-0.123	0.326	0.906	0.817	0.956	0.120	0.148	0.220	-0.188	-0.128	0.326	0.896	0.818	0.956
10	0.189	0.182	0.221	-0.161	-0.118	0.285	0.781	0.742	0.945	0.188	0.182	0.221	-0.071	-0.062	0.285	0.773	0.734	0.945
11	0.147	0.175	0.25	-0.14	-0.158	0.321	0.871	0.803	0.942	0.145	0.175	0.250	-0.051	-0.103	0.321	0.868	0.797	0.942
12	0.154	0.172	0.214	-0.174	-0.128	0.307	0.839	0.802	0.949	0.154	0.170	0.214	-0.092	-0.063	0.307	0.838	0.789	0.949
13	0.129	0.149	0.217	-0.185	-0.162	0.326	0.905	0.851	0.954	0.121	0.144	0.217	-0.084	-0.089	0.326	0.897	0.843	0.954
14	0.137	0.151	0.206	-0.183	-0.131	0.332	0.873	0.811	0.959	0.135	0.151	0.206	-0.112	-0.078	0.332	0.882	0.806	0.959
15	0.144	0.15	0.258	-0.177	-0.126	0.304	0.879	0.821	0.935	0.142	0.148	0.258	-0.086	-0.063	0.304	0.873	0.809	0.935
16	0.136	0.122	0.202	-0.146	-0.123	0.277	0.887	0.901	0.953	0.134	0.118	0.202	-0.075	-0.082	0.277	0.891	0.903	0.953
17	0.121	0.096	0.219	-0.169	-0.108	0.325	0.897	0.925	0.951	0.121	0.093	0.219	-0.087	-0.062	0.325	0.893	0.919	0.951
18	0.184	0.157	0.263	-0.176	-0.156	0.332	0.799	0.810	0.930	0.183	0.155	0.263	-0.102	-0.112	0.332	0.806	0.821	0.930
19	0.127	0.195	0.335	-0.163	-0.214	0.425	0.886	0.748	0.934	0.125	0.186	0.335	-0.089	-0.130	0.425	0.894	0.704	0.934
20	0.127	0.167	0.291	-0.107	-0.102	0.359	0.907	0.783	0.922	0.127	0.167	0.291	-0.107	-0.102	0.359	0.903	0.776	0.922
MEAN	0.151	0.157	0.253	-0.166	-0.144	0.311	0.862	0.822	0.933	0.147	0.154	0.253	-0.128	-0.117	0.311	0.865	0.818	0.933
SD.	0.038	0.028	0.084	0.021	0.027	0.071	0.053	0.054	0.061	0.034	0.028	0.084	0.050	0.041	0.071	0.042	0.059	0.061
Min	0.121	0.096	0.125	-0.205	-0.214	0.070	0.716	0.723	0.682	0.120	0.093	0.125	-0.218	-0.196	0.070	0.773	0.704	0.682
Max	0.262	0.197	0.557	-0.107	-0.102	0.425	0.914	0.925	0.982	0.243	0.193	0.557	-0.051	-0.062	0.425	0.913	0.921	0.982

ตารางภาคผนวกที่ 3 ค่าดัชนี RMSE BIAS และสัมประสิทธิ์สหสัมพันธ์ของเงื่อนไขมีความสัมพันธ์ระหว่างมิติความสามารถระดับกลางในโครงสร้างที่ไม่ซับซ้อนของวิธี M และ วิธี NOP สำหรับการแปลงคะแนนจากระดับชั้นที่ 1 ไป 2 จากการทำซ้ำ 20 ครั้ง

ลำดับในการทำซ้ำ	L_12 CORR 0.5 APSS																	
	APSS_M									APSS_NOP								
	RMSE			BIAS_TRUE			Correlation Coefficient			RMSE			BIAS_TRUE			Correlation Coefficient		
	a1	a2	d	a1	a2	d	a1	a2	d	a1	a1	d	a1	a1	d	a1	a1	d
1	0.143	0.108	0.194	-0.063	-0.093	0.35	0.818	0.884	0.965	0.141	0.103	0.194	-0.071	-0.080	0.350	0.862	0.876	0.965
2	0.146	0.112	0.234	-0.088	-0.087	0.384	0.853	0.884	0.962	0.144	0.109	0.234	-0.078	-0.075	0.384	0.844	0.871	0.962
3	0.138	0.126	0.248	-0.062	-0.103	0.367	0.869	0.877	0.946	0.138	0.123	0.248	-0.063	-0.089	0.367	0.866	0.859	0.946
4	0.128	0.141	0.21	-0.051	-0.094	0.355	0.892	0.792	0.955	0.128	0.136	0.210	-0.050	-0.080	0.355	0.890	0.777	0.955
5	0.107	0.133	0.138	-0.083	-0.082	0.344	0.906	0.844	0.975	0.107	0.133	0.138	-0.082	-0.082	0.344	0.902	0.837	0.975
6	0.106	0.101	0.205	-0.061	-0.084	0.395	0.904	0.909	0.974	0.105	0.100	0.205	-0.065	-0.082	0.395	0.906	0.908	0.974
7	0.135	0.116	0.177	-0.075	-0.101	0.346	0.883	0.902	0.977	0.135	0.111	0.177	-0.069	-0.086	0.346	0.882	0.900	0.977
8	0.104	0.153	0.204	-0.092	-0.118	0.384	0.909	0.773	0.966	0.104	0.150	0.204	-0.093	-0.106	0.384	0.909	0.765	0.966
9	0.123	0.149	0.232	-0.085	-0.095	0.36	0.876	0.785	0.963	0.122	0.144	0.232	-0.081	-0.081	0.360	0.872	0.782	0.963
10	0.159	0.115	0.202	-0.063	-0.081	0.304	0.795	0.900	0.955	0.159	0.114	0.202	-0.063	-0.075	0.304	0.794	0.894	0.955
11	0.131	0.092	0.237	-0.083	-0.082	0.362	0.858	0.925	0.967	0.130	0.092	0.237	-0.085	-0.079	0.362	0.860	0.924	0.967
12	0.144	0.087	0.184	-0.094	-0.073	0.365	0.817	0.936	0.972	0.143	0.087	0.184	-0.085	-0.074	0.365	0.809	0.936	0.972
13	0.203	0.117	0.262	-0.116	-0.097	0.384	0.705	0.888	0.951	0.195	0.107	0.262	-0.091	-0.078	0.384	0.684	0.883	0.951
14	0.189	0.136	0.266	-0.109	-0.097	0.394	0.803	0.837	0.942	0.186	0.135	0.266	-0.092	-0.093	0.394	0.781	0.832	0.942
15	0.09	0.102	0.205	-0.08	-0.077	0.394	0.919	0.916	0.970	0.090	0.098	0.205	-0.080	-0.064	0.394	0.919	0.920	0.970
16	0.136	0.133	0.214	-0.075	-0.135	0.399	0.862	0.872	0.975	0.136	0.124	0.214	-0.076	-0.112	0.399	0.862	0.847	0.975
17	0.145	0.139	0.26	-0.063	-0.101	0.361	0.830	0.835	0.942	0.143	0.134	0.260	-0.072	-0.087	0.361	0.844	0.820	0.942
18	0.111	0.106	0.178	-0.075	-0.067	0.340	0.899	0.905	0.974	0.111	0.105	0.178	-0.073	-0.061	0.340	0.898	0.906	0.974
19	0.117	0.104	0.172	-0.085	-0.069	0.358	0.885	0.911	0.971	0.117	0.104	0.172	-0.084	-0.071	0.358	0.885	0.912	0.971
20	0.14	0.12	0.247	-0.095	-0.082	0.368	0.877	0.858	0.962	0.135	0.118	0.247	-0.079	-0.074	0.368	0.863	0.852	0.962
MEAN	0.135	0.120	0.213	-0.080	-0.091	0.366	0.858	0.872	0.963	0.133	0.116	0.213	-0.077	-0.081	0.366	0.857	0.865	0.963
SD.	0.027	0.019	0.034	0.017	0.016	0.023	0.051	0.047	0.011	0.026	0.018	0.034	0.011	0.012	0.023	0.055	0.051	0.011
Min	0.090	0.087	0.138	-0.116	-0.135	0.304	0.705	0.773	0.942	0.090	0.087	0.138	-0.093	-0.112	0.304	0.684	0.765	0.942
Max	0.203	0.153	0.266	-0.051	-0.067	0.399	0.919	0.936	0.977	0.195	0.150	0.266	-0.050	-0.061	0.399	0.919	0.936	0.977

จุฬาลงกรณ์มหาวิทยาลัย

ตารางภาคผนวกที่ 4 ค่าดัชนี RMSE BIAS และสัมประสิทธิ์สหสัมพันธ์ของเงื่อนไขมีความสัมพันธ์ระหว่างมิติความสามารถระดับกลางในโครงสร้างที่ซับซ้อนของวิธี M และ วิธี NOP สำหรับการแปลงคะแนนจากระดับชั้นที่ 1 ไป 2 จากการทำซ้ำ 20 ครั้ง

ลำดับในการทำซ้ำ	L_12 Test CORR 0.5 MS																	
	MS_M									MS_NOP								
	RMSE			BIAS_TRUE			Correlation Coefficient			RMSE			BIAS_TRUE			Correlation Coefficient		
	a1	a2	d	a1	a2	d	a1	a2	d	a1	a1	d	a1	a1	d	a1	a1	d
1	0.132	0.112	0.199	-0.116	-0.116	0.481	0.789	0.851	0.968	0.128	0.103	0.199	-0.107	-0.101	0.481	0.792	0.851	0.967
2	0.124	0.127	0.170	-0.115	-0.112	0.427	0.838	0.805	0.975	0.123	0.114	0.170	-0.110	-0.094	0.427	0.834	0.813	0.975
3	0.147	0.129	0.171	-0.145	-0.101	0.434	0.830	0.762	0.974	0.130	0.129	0.171	-0.118	-0.099	0.434	0.817	0.756	0.974
4	0.127	0.099	0.185	-0.099	-0.126	0.452	0.807	0.889	0.971	0.127	0.086	0.185	-0.099	-0.109	0.452	0.807	0.892	0.971
5	0.132	0.137	0.212	-0.107	-0.105	0.437	0.818	0.777	0.969	0.125	0.137	0.212	-0.094	-0.105	0.437	0.812	0.777	0.969
6	0.112	0.094	0.181	-0.099	-0.112	0.493	0.872	0.874	0.974	0.112	0.091	0.181	-0.100	-0.105	0.493	0.872	0.872	0.974
7	0.126	0.140	0.200	-0.114	-0.105	0.470	0.789	0.745	0.969	0.126	0.140	0.200	-0.114	-0.105	0.470	0.789	0.745	0.969
8	0.134	0.140	0.178	-0.135	-0.096	0.464	0.799	0.754	0.977	0.128	0.121	0.178	-0.121	-0.075	0.464	0.791	0.782	0.977
9	0.142	0.142	0.218	-0.141	-0.116	0.458	0.777	0.727	0.969	0.134	0.134	0.218	-0.046	-0.049	0.458	0.754	0.722	0.969
10	0.149	0.122	0.172	-0.112	-0.117	0.495	0.720	0.784	0.974	0.146	0.118	0.172	-0.025	-0.054	0.495	0.720	0.777	0.974
11	0.123	0.145	0.182	-0.089	-0.138	0.461	0.851	0.722	0.966	0.116	0.143	0.182	0.002	-0.077	0.461	0.851	0.706	0.966
12	0.114	0.144	0.172	-0.091	-0.122	0.459	0.847	0.747	0.972	0.112	0.141	0.172	-0.006	-0.058	0.459	0.841	0.732	0.972
13	0.133	0.130	0.208	-0.095	-0.121	0.475	0.814	0.797	0.959	0.132	0.119	0.208	-0.091	-0.104	0.475	0.814	0.765	0.959
14	0.199	0.141	0.171	-0.154	-0.070	0.453	0.659	0.713	0.973	0.172	0.141	0.171	-0.036	-0.017	0.453	0.638	0.719	0.973
15	0.152	0.134	0.179	-0.143	-0.110	0.468	0.735	0.766	0.977	0.149	0.117	0.179	-0.053	-0.037	0.468	0.725	0.798	0.977
16	0.108	0.156	0.152	-0.082	-0.144	0.437	0.860	0.783	0.981	0.107	0.133	0.152	-0.004	-0.059	0.437	0.862	0.747	0.981
17	0.148	0.119	0.191	-0.106	-0.151	0.466	0.739	0.811	0.969	0.147	0.096	0.191	-0.030	-0.075	0.466	0.753	0.831	0.969
18	0.133	0.104	0.196	-0.106	-0.108	0.462	0.805	0.863	0.970	0.130	0.099	0.196	-0.018	-0.046	0.462	0.806	0.846	0.970
19	0.115	0.145	0.170	-0.141	-0.102	0.441	0.850	0.751	0.970	0.114	0.143	0.170	-0.056	-0.041	0.441	0.842	0.738	0.970
20	0.125	0.134	0.114	-0.105	-0.123	0.477	0.826	0.783	0.988	0.125	0.128	0.114	-0.023	-0.056	0.477	0.830	0.766	0.988
MEAN	0.134	0.130	0.181	-0.115	-0.115	0.461	0.801	0.785	0.972	0.129	0.122	0.181	-0.062	-0.073	0.461	0.797	0.782	0.972
SD.	0.020	0.017	0.023	0.021	0.018	0.019	0.054	0.051	0.006	0.015	0.018	0.023	0.043	0.028	0.019	0.056	0.053	0.006
Min	0.108	0.094	0.114	-0.154	-0.151	0.427	0.659	0.713	0.959	0.107	0.086	0.114	-0.121	-0.109	0.427	0.638	0.706	0.959
Max	0.199	0.156	0.218	-0.082	-0.070	0.495	0.872	0.889	0.988	0.172	0.143	0.218	0.002	-0.017	0.495	0.872	0.892	0.988

ตารางภาคผนวกที่ 5 ค่าดัชนี RMSE BIAS และสัมประสิทธิ์สหสัมพันธ์ของเงื่อนไขมีความสัมพันธ์ระหว่างมิติความสามารถระดับสูงในโครงสร้างที่ไม่ซับซ้อนของวิธี M และ วิธี NOP สำหรับการแปลงคะแนนจากระดับชั้นที่ 1 ไป 2 จากการทำซ้ำ 20 ครั้ง

ลำดับในการทำซ้ำ	L_12 CORR 0.8 APSS																	
	APSS_M									APSS_NOP								
	RMSE			BIAS_TRUE			Correlation Coefficient			RMSE			BIAS_TRUE			Correlation Coefficient		
	a1	a2	d	a1	a2	d	a1	a2	d	a1	a1	d	a1	a1	d	a1	a1	d
1	0.191	0.123	0.180	-0.057	-0.017	0.394	0.537	0.789	0.969	0.183	0.115	0.180	-0.042	-0.006	0.394	0.496	0.770	0.969
2	0.166	0.133	0.188	-0.095	-0.012	0.395	0.657	0.785	0.968	0.164	0.123	0.188	-0.087	0.002	0.395	0.638	0.779	0.968
3	0.184	0.143	0.266	-0.130	-0.055	0.488	0.727	0.766	0.964	0.173	0.128	0.266	-0.105	-0.040	0.488	0.649	0.780	0.964
4	0.157	0.161	0.213	-0.053	-0.085	0.427	0.663	0.646	0.968	0.142	0.147	0.213	-0.035	-0.065	0.427	0.641	0.604	0.968
5	0.116	0.143	0.162	-0.058	-0.067	0.432	0.745	0.720	0.972	0.114	0.131	0.162	-0.053	-0.051	0.432	0.734	0.712	0.972
6	0.225	0.158	0.303	-0.075	-0.098	0.509	0.428	0.709	0.951	0.206	0.142	0.303	-0.052	-0.079	0.509	0.384	0.688	0.951
7	0.206	0.124	0.223	-0.036	-0.044	0.394	0.510	0.778	0.960	0.201	0.118	0.223	-0.025	-0.034	0.394	0.494	0.767	0.960
8	0.148	0.118	0.210	-0.052	-0.047	0.428	0.708	0.808	0.971	0.134	0.112	0.210	-0.034	-0.036	0.428	0.703	0.802	0.971
9	0.148	0.126	0.219	-0.063	-0.020	0.433	0.705	0.750	0.964	0.135	0.097	0.219	-0.047	-0.001	0.433	0.705	0.802	0.964
10	0.140	0.148	0.138	-0.046	-0.045	0.386	0.716	0.675	0.984	0.138	0.144	0.138	-0.041	-0.037	0.386	0.718	0.662	0.984
11	0.141	0.188	0.212	-0.068	-0.029	0.417	0.690	0.575	0.955	0.139	0.179	0.212	-0.062	-0.013	0.417	0.688	0.564	0.955
12	0.147	0.198	0.142	-0.056	-0.060	0.405	0.686	0.686	0.981	0.141	0.177	0.142	-0.046	-0.035	0.405	0.682	0.461	0.981
13	0.115	0.170	0.166	-0.071	-0.025	0.402	0.832	0.627	0.979	0.098	0.156	0.166	-0.052	-0.007	0.402	0.827	0.627	0.979
14	0.138	0.120	0.172	-0.037	-0.061	0.429	0.763	0.807	0.975	0.136	0.114	0.172	-0.031	-0.050	0.429	0.754	0.799	0.975
15	0.137	0.181	0.184	-0.002	-0.097	0.375	0.675	0.646	0.977	0.122	0.161	0.184	0.010	-0.068	0.375	0.699	0.655	0.977
16	0.124	0.164	0.228	-0.049	-0.023	0.418	0.778	0.669	0.954	0.115	0.162	0.228	-0.038	-0.018	0.418	0.753	0.661	0.954
17	0.215	0.175	0.199	-0.060	-0.055	0.391	0.559	0.608	0.966	0.200	0.152	0.199	-0.034	-0.033	0.391	0.517	0.650	0.966
18	0.138	0.129	0.135	-0.062	-0.041	0.394	0.714	0.780	0.984	0.133	0.122	0.135	-0.053	-0.030	0.394	0.677	0.788	0.984
19	0.127	0.124	0.175	-0.074	-0.024	0.386	0.828	0.769	0.973	0.123	0.110	0.175	-0.063	-0.012	0.386	0.807	0.807	0.973
20	0.168	0.106	0.209	-0.037	-0.051	0.377	0.638	0.827	0.959	0.163	0.096	0.209	-0.027	-0.039	0.377	0.623	0.819	0.959
MEAN	0.157	0.147	0.196	-0.059	-0.048	0.414	0.678	0.721	0.969	0.148	0.134	0.196	-0.046	-0.033	0.414	0.659	0.710	0.969
SD.	0.032	0.026	0.042	0.025	0.025	0.034	0.103	0.075	0.010	0.031	0.025	0.042	0.024	0.023	0.034	0.111	0.098	0.010
Min	0.115	0.106	0.135	-0.130	-0.098	0.375	0.428	0.575	0.951	0.098	0.096	0.135	-0.105	-0.079	0.375	0.384	0.461	0.951
Max	0.225	0.198	0.303	-0.002	-0.012	0.509	0.832	0.827	0.984	0.206	0.179	0.303	0.010	0.002	0.509	0.827	0.819	0.984

ตารางภาคผนวกที่ 6 ค่าดัชนี RMSE BIAS และสัมประสิทธิ์สหสัมพันธ์ของเงื่อนไขมีความสัมพันธ์ระหว่างมิติความสามารถระดับสูงในโครงสร้างที่ซับซ้อนของวิธี M และ วิธี NOP สำหรับการแปลงคะแนนจากระดับชั้นที่ 1 ไป 2 จากการทำซ้ำ 20 ครั้ง

ลำดับในการทำซ้ำ	L_12 CORR 0.8 MS																	
	MS_M									MS_NOP								
	RMSE			BIAS_TRUE			Correlation Coefficient			RMSE			BIAS_TRUE			Correlation Coefficient		
	a1	a2	d	a1	a2	d	a1	a2	d	a1	a1	d	a1	a1	d	a1	a1	d
1	0.186	0.213	0.138	-0.102	-0.048	0.472	0.590	0.625	0.979	0.135	0.196	0.138	-0.072	-0.030	0.472	0.745	0.389	0.979
2	0.17	0.18	0.122	-0.077	-0.1	0.477	0.633	0.496	0.985	0.146	0.146	0.122	-0.058	-0.075	0.477	0.668	0.517	0.985
3	0.221	0.262	0.193	-0.124	-0.077	0.47	0.378	0.276	0.968	0.184	0.152	0.193	-0.097	-0.014	0.470	0.495	0.629	0.968
4	0.196	0.223	0.181	0.009	-0.271	0.51	0.424	0.334	0.984	0.138	0.164	0.181	0.030	-0.206	0.510	0.670	0.425	0.984
5	0.254	0.205	0.228	-0.143	-0.058	0.504	0.267	0.567	0.959	0.232	0.164	0.228	-0.122	-0.027	0.504	0.295	0.604	0.959
6	0.164	0.251	0.16	-0.048	-0.117	0.463	0.721	0.224	0.974	0.137	0.210	0.160	-0.021	-0.098	0.463	0.720	0.537	0.974
7	0.2	0.241	0.249	-0.057	-0.124	0.47	0.637	0.332	0.939	0.168	0.229	0.249	-0.028	-0.108	0.470	0.632	0.332	0.939
8	0.186	0.179	0.164	-0.098	-0.004	0.451	0.483	0.684	0.973	0.131	0.158	0.164	-0.067	0.015	0.451	0.674	0.760	0.973
9	0.224	0.21	0.185	-0.166	-0.034	0.14	0.267	0.374	0.969	0.176	0.171	0.185	-0.049	0.042	0.140	0.374	0.509	0.969
10	0.186	0.179	0.164	-0.098	-0.004	0.451	0.483	0.684	0.973	0.131	0.158	0.164	0.013	0.069	0.451	0.674	0.760	0.973
11	0.144	0.213	0.19	-0.04	-0.125	0.445	0.667	0.439	0.967	0.136	0.169	0.190	0.048	-0.032	0.445	0.668	0.533	0.967
12	0.19	0.142	0.11	-0.127	-0.067	0.494	0.549	0.712	0.987	0.170	0.110	0.110	-0.026	0.007	0.494	0.572	0.742	0.987
13	0.161	0.261	0.208	-0.115	-0.065	0.489	0.596	0.460	0.960	0.148	0.154	0.208	-0.102	0.010	0.489	0.611	0.694	0.960
14	0.19	0.21	0.207	-0.07	-0.113	0.487	0.388	0.396	0.968	0.153	0.137	0.207	0.028	-0.018	0.487	0.536	0.691	0.968
15	0.203	0.194	0.159	-0.09	-0.057	0.49	0.462	0.544	0.974	0.165	0.183	0.159	0.019	0.010	0.490	0.502	0.547	0.974
16	0.227	0.197	0.422	-0.079	0.053	-0.682	0.480	0.692	0.867	0.209	0.197	0.422	0.023	0.108	-0.682	0.491	0.690	0.867
17	0.186	0.171	0.163	-0.096	-0.082	0.134	0.656	0.444	0.978	0.169	0.147	0.163	0.009	-0.017	0.134	0.607	0.588	0.978
18	0.206	0.19	0.128	-0.102	-0.058	0.483	0.436	0.539	0.984	0.170	0.146	0.128	-0.001	0.028	0.483	0.608	0.622	0.984
19	0.182	0.192	0.218	-0.044	-0.137	0.471	0.404	0.525	0.962	0.145	0.144	0.218	0.051	-0.039	0.471	0.614	0.573	0.962
20	0.209	0.174	0.156	-0.151	-0.032	0.113	0.452	0.548	0.965	0.166	0.130	0.156	-0.031	0.048	0.113	0.617	0.616	0.965
MEAN	0.194	0.204	0.187	-0.091	-0.076	0.367	0.499	0.495	0.966	0.160	0.16	0.19	-0.02	-0.02	0.37	0.59	0.59	0.97
SD.	0.025	0.032	0.066	0.042	0.066	0.278	0.129	0.144	0.026	0.026	0.03	0.07	0.05	0.07	0.28	0.11	0.12	0.03
Min	0.144	0.142	0.110	-0.166	-0.271	-0.682	0.267	0.224	0.867	0.131	0.110	0.110	-0.122	-0.206	-0.682	0.295	0.332	0.867
Max	0.254	0.262	0.422	0.009	0.053	0.510	0.721	0.712	0.987	0.232	0.229	0.422	0.051	0.108	0.510	0.745	0.760	0.987

ตารางภาคผนวกที่ 7 ค่าดัชนี RMSE BIAS และสัมประสิทธิ์สหสัมพันธ์ของเงื่อนไขไม่มีความสัมพันธ์ระหว่างมิติความสามารถในโครงสร้างที่ไม่ซับซ้อนของวิธี M และ วิธี NOP สำหรับการแปลงคะแนนจากระดับชั้นที่ 2 ไป 3 จากการทำซ้ำ 20 ครั้ง

ลำดับในการทำซ้ำ	L _23 CORR 0.0 APSS																	
	APSS_M									APSS_NOP								
	RMSE			BIAS_TRUE			Correlation Coefficient			RMSE			BIAS_TRUE			Correlation Coefficient		
	a1	a2	d	a1	a2	d	a1	a2	d	a1	a1	d	a1	a1	d	a1	a1	d
1	0.086	0.098	0.190	-0.176	-0.143	0.852	0.965	0.965	0.962	0.086	0.095	0.190	-0.175	-0.155	0.852	0.965	0.965	0.962
2	0.144	0.114	0.237	-0.199	-0.182	0.858	0.921	0.940	0.946	0.139	0.114	0.237	-0.175	-0.184	0.858	0.917	0.940	0.946
3	0.152	0.139	0.251	-0.192	-0.160	0.807	0.918	0.920	0.937	0.147	0.138	0.251	-0.169	-0.164	0.807	0.915	0.921	0.937
4	0.143	0.135	0.316	-0.190	-0.176	0.824	0.932	0.920	0.916	0.135	0.134	0.316	-0.160	-0.166	0.824	0.927	0.920	0.916
5	0.138	0.136	0.206	-0.140	-0.153	0.791	0.934	0.924	0.953	0.138	0.126	0.206	-0.140	-0.179	0.791	0.934	0.935	0.953
6	0.111	0.157	0.209	-0.177	-0.158	0.861	0.941	0.903	0.962	0.111	0.156	0.209	-0.180	-0.170	0.861	0.942	0.908	0.962
7	0.149	0.102	0.199	-0.148	-0.164	0.809	0.914	0.956	0.958	0.148	0.102	0.199	-0.151	-0.168	0.809	0.915	0.957	0.958
8	0.126	0.126	0.214	-0.191	-0.188	0.880	0.933	0.925	0.968	0.125	0.126	0.214	-0.181	-0.191	0.880	0.930	0.926	0.968
9	0.119	0.131	0.220	-0.134	-0.177	0.807	0.948	0.915	0.958	0.118	0.131	0.220	-0.142	-0.182	0.807	0.950	0.917	0.958
10	0.123	0.127	0.216	-0.173	-0.179	0.857	0.935	0.924	0.956	0.123	0.126	0.216	-0.165	-0.174	0.857	0.935	0.924	0.956
11	0.153	0.115	0.287	-0.148	-0.175	0.846	0.915	0.939	0.919	0.153	0.115	0.287	-0.144	-0.176	0.846	0.915	0.939	0.919
12	0.137	0.107	0.243	-0.151	-0.170	0.820	0.928	0.953	0.959	0.136	0.107	0.243	-0.159	-0.173	0.820	0.930	0.954	0.959
13	0.123	0.139	0.257	-0.147	-0.184	0.795	0.945	0.919	0.931	0.123	0.138	0.257	-0.145	-0.174	0.795	0.945	0.917	0.931
14	0.131	0.144	0.289	-0.167	-0.176	0.804	0.933	0.909	0.932	0.130	0.144	0.289	-0.156	-0.176	0.804	0.932	0.909	0.932
15	0.192	0.124	0.324	-0.100	-0.190	0.695	0.888	0.918	0.913	0.191	0.123	0.324	-0.106	-0.181	0.695	0.889	0.917	0.913
16	0.173	0.123	0.280	-0.132	-0.201	0.733	0.912	0.929	0.921	0.172	0.122	0.280	-0.119	-0.190	0.733	0.911	0.926	0.921
17	0.116	0.137	0.189	-0.179	-0.145	0.835	0.943	0.926	0.968	0.113	0.136	0.189	-0.165	-0.153	0.835	0.942	0.927	0.968
18	0.105	0.116	0.212	-0.179	-0.169	0.876	0.955	0.942	0.959	0.104	0.115	0.212	-0.168	-0.177	0.876	0.954	0.943	0.959
19	0.103	0.101	0.220	-0.155	-0.185	0.856	0.960	0.952	0.957	0.102	0.101	0.220	-0.163	-0.187	0.856	0.961	0.952	0.957
20	0.108	0.109	0.264	-0.156	-0.156	0.823	0.950	0.952	0.938	0.107	0.109	0.264	-0.148	-0.158	0.823	0.949	0.953	0.938
MEAN	0.132	0.124	0.241	-0.162	-0.172	0.821	0.933	0.932	0.946	0.130	0.123	0.241	-0.156	-0.174	0.821	0.933	0.933	0.946
SD.	0.025	0.016	0.041	0.025	0.015	0.046	0.019	0.016	0.018	0.025	0.016	0.041	0.019	0.011	0.046	0.019	0.017	0.018
Min	0.086	0.098	0.189	-0.199	-0.201	0.695	0.888	0.903	0.913	0.086	0.095	0.189	-0.181	-0.191	0.695	0.889	0.908	0.913
Max	0.192	0.157	0.324	-0.100	-0.143	0.880	0.965	0.965	0.968	0.191	0.156	0.324	-0.106	-0.153	0.880	0.965	0.965	0.968

ตารางภาคผนวกที่ 8 ค่าดัชนี RMSE BIAS และสัมประสิทธิ์สหสัมพันธ์ของเงื่อนไขไม่มีความสัมพันธ์ระหว่างมิติความสามารถในโครงสร้างที่ซับซ้อนของวิธี M และ วิธี NOP สำหรับการแปลงคะแนนจากระดับชั้นที่ 2 ไป 3 จากการทำซ้ำ 20 ครั้ง

ลำดับในการทำซ้ำ	L_23 CORR 0.0 MS																	
	MS_M									MS_NOP								
	RMSE			BIAS_TRUE			Correlation Coefficient			RMSE			BIAS_TRUE			Correlation Coefficient		
	a1	a2	d	a1	a2	d	a1	a2	d	a1	a1	d	a1	a1	d	a1	a1	d
1	0.091	0.090	0.172	-0.228	-0.190	1.042	0.934	0.934	0.974	0.091	0.089	0.172	-0.230	-0.193	1.042	0.935	0.935	0.974
2	0.207	0.167	0.276	-0.297	-0.227	1.059	0.754	0.755	0.931	0.205	0.151	0.276	-0.280	-0.248	1.059	0.728	0.826	0.931
3	0.107	0.115	0.173	-0.226	-0.193	1.027	0.916	0.910	0.977	0.105	0.114	0.173	-0.217	-0.196	1.027	0.916	0.911	0.977
4	0.130	0.133	0.231	-0.184	-0.205	1.024	0.901	0.885	0.961	0.126	0.133	0.231	-0.196	-0.203	1.024	0.905	0.884	0.961
5	0.080	0.125	0.197	-0.202	-0.206	1.015	0.955	0.903	0.970	0.079	0.125	0.197	-0.200	-0.204	1.015	0.955	0.903	0.970
6	0.098	0.118	0.239	-0.208	-0.192	1.048	0.936	0.881	0.949	0.096	0.118	0.239	-0.198	-0.190	1.048	0.936	0.881	0.949
7	0.108	0.123	0.278	-0.198	-0.198	1.017	0.920	0.888	0.924	0.108	0.123	0.278	-0.196	-0.199	1.017	0.919	0.888	0.924
8	0.131	0.128	0.178	-0.227	-0.209	1.054	0.872	0.879	0.971	0.127	0.127	0.178	-0.214	-0.205	1.054	0.871	0.877	0.971
9	0.124	0.148	0.234	-0.213	-0.201	1.002	0.905	0.838	0.943	0.124	0.148	0.234	-0.212	-0.204	1.002	0.905	0.840	0.943
10	0.105	0.108	0.177	-0.213	-0.208	1.046	0.922	0.906	0.973	0.105	0.108	0.177	-0.213	-0.208	1.046	0.922	0.907	0.973
11	0.099	0.106	0.178	-0.190	-0.208	1.036	0.933	0.916	0.967	0.098	0.106	0.178	-0.186	-0.209	1.036	0.933	0.916	0.967
12	0.170	0.134	0.284	-0.222	-0.222	1.020	0.800	0.882	0.958	0.164	0.125	0.284	-0.203	-0.199	1.020	0.789	0.874	0.958
13	0.127	0.105	0.200	-0.207	-0.203	1.013	0.880	0.912	0.968	0.126	0.104	0.200	-0.199	-0.198	1.013	0.879	0.909	0.968
14	0.114	0.128	0.166	-0.213	-0.197	1.021	0.898	0.878	0.977	0.114	0.128	0.166	-0.212	-0.193	1.021	0.897	0.878	0.977
15	0.097	0.116	0.232	-0.202	-0.212	0.998	0.941	0.907	0.960	0.097	0.114	0.232	-0.205	-0.221	0.998	0.942	0.911	0.960
16	0.130	0.093	0.213	-0.223	-0.218	1.051	0.878	0.926	0.961	0.130	0.093	0.213	-0.221	-0.221	1.051	0.876	0.928	0.961
17	0.114	0.116	0.194	-0.205	-0.225	1.036	0.892	0.886	0.967	0.114	0.115	0.194	-0.203	-0.219	1.036	0.890	0.885	0.967
18	0.117	0.118	0.187	-0.196	-0.203	1.055	0.901	0.897	0.977	0.116	0.117	0.187	-0.202	-0.209	1.055	0.904	0.899	0.977
19	0.132	0.107	0.196	-0.183	-0.211	1.013	0.909	0.910	0.966	0.131	0.105	0.196	-0.191	-0.205	1.013	0.911	0.908	0.966
20	0.110	0.141	0.171	-0.176	-0.213	1.002	0.930	0.858	0.970	0.110	0.140	0.171	-0.180	-0.222	1.002	0.929	0.863	0.970
MEAN	0.120	0.121	0.209	-0.211	-0.207	1.029	0.899	0.888	0.962	0.118	0.119	0.209	-0.208	-0.207	1.029	0.897	0.891	0.962
SD.	0.028	0.018	0.038	0.025	0.011	0.019	0.048	0.039	0.015	0.027	0.016	0.038	0.021	0.014	0.019	0.053	0.027	0.015
Min	0.080	0.090	0.166	-0.297	-0.227	0.998	0.754	0.755	0.924	0.079	0.089	0.166	-0.280	-0.248	0.998	0.728	0.826	0.924
Max	0.207	0.167	0.284	-0.176	-0.190	1.059	0.955	0.934	0.977	0.205	0.151	0.284	-0.180	-0.190	1.059	0.955	0.935	0.977

ตารางภาคผนวกที่ 9 ค่าดัชนี RMSE BIAS และสัมประสิทธิ์สหสัมพันธ์ของเงื่อนไขความสัมพันธ์ระหว่างมิติความสามารถระดับกลางในโครงสร้างที่ไม่ซับซ้อนของวิธี M และ วิธี NOP สำหรับการแปลงคะแนนจากระดับชั้นที่ 2 ไป 3 จากการทำซ้ำ 20 ครั้ง

ลำดับในการทำซ้ำ	L_23 Test CORR 0.5 APSS																	
	APSS_M									APSS_NOP								
	RMSE			BIAS_TRUE			Correlation Coefficient			RMSE			BIAS_TRUE			Correlation Coefficient		
	a1	a2	d	a1	a2	d	a1	a2	d	a1	a1	d	a1	a1	d	a1	a1	d
1	0.094	0.121	0.206	-0.083	-0.081	0.949	0.924	0.893	0.961	0.094	0.118	0.206	-0.085	-0.089	0.949	0.924	0.893	0.961
2	0.097	0.132	0.206	-0.086	-0.105	0.959	0.936	0.874	0.972	0.095	0.131	0.206	-0.079	-0.099	0.959	0.936	0.872	0.972
3	0.090	0.097	0.199	-0.095	-0.091	0.977	0.938	0.927	0.966	0.090	0.097	0.199	-0.091	-0.095	0.977	0.938	0.929	0.966
4	0.054	0.060	0.260	-0.078	-0.078	1.006	0.982	0.972	0.914	0.052	0.060	0.260	-0.072	-0.080	1.006	0.981	0.973	0.914
5	0.155	0.101	0.255	-0.114	-0.117	0.973	0.835	0.914	0.945	0.148	0.093	0.255	-0.095	-0.102	0.973	0.818	0.915	0.945
6	0.105	0.126	0.180	-0.097	-0.081	0.956	0.921	0.882	0.886	0.098	0.125	0.180	-0.081	-0.086	0.956	0.919	0.886	0.886
7	0.100	0.116	0.191	-0.104	-0.096	0.957	0.929	0.889	0.973	0.092	0.114	0.191	-0.089	-0.089	0.957	0.922	0.886	0.973
8	0.103	0.121	0.208	-0.094	-0.115	0.958	0.919	0.903	0.967	0.097	0.116	0.208	-0.081	-0.100	0.958	0.924	0.893	0.967
9	0.117	0.133	0.205	-0.115	-0.085	0.937	0.903	0.880	0.974	0.112	0.133	0.205	-0.101	-0.084	0.937	0.895	0.880	0.974
10	0.118	0.096	0.203	-0.113	-0.092	0.945	0.899	0.932	0.972	0.118	0.096	0.203	-0.113	-0.092	0.945	0.899	0.932	0.972
11	0.116	0.089	0.200	-0.118	-0.105	0.979	0.897	0.935	0.970	0.116	0.088	0.200	-0.115	-0.104	0.979	0.894	0.935	0.970
12	0.207	0.182	0.430	-0.113	-0.122	0.968	0.676	0.767	0.732	0.207	0.179	0.430	-0.112	-0.110	0.968	0.675	0.745	0.732
13	0.133	0.108	0.233	-0.090	-0.096	0.951	0.870	0.910	0.955	0.132	0.104	0.233	-0.085	-0.084	0.951	0.866	0.909	0.955
14	0.081	0.144	0.214	-0.104	-0.097	0.984	0.946	0.843	0.974	0.079	0.144	0.214	-0.097	-0.099	0.984	0.943	0.846	0.974
15	0.090	0.154	0.190	-0.081	-0.118	0.977	0.947	0.810	0.978	0.089	0.148	0.190	-0.076	-0.102	0.977	0.946	0.795	0.978
16	0.120	0.120	0.201	-0.101	-0.108	0.943	0.907	0.881	0.972	0.118	0.116	0.201	-0.093	-0.096	0.943	0.904	0.879	0.972
17	0.120	0.100	0.172	-0.086	-0.090	0.970	0.886	0.912	0.974	0.119	0.100	0.172	-0.082	-0.091	0.970	0.887	0.912	0.974
18	0.115	0.131	0.278	-0.107	-0.088	0.973	0.890	0.862	0.944	0.114	0.130	0.278	-0.100	-0.094	0.973	0.887	0.866	0.944
19	0.098	0.135	0.249	-0.089	-0.105	0.958	0.924	0.871	0.962	0.098	0.134	0.249	-0.089	-0.098	0.958	0.924	0.867	0.962
20	0.086	0.121	0.165	-0.087	-0.091	0.949	0.945	0.874	0.972	0.085	0.119	0.165	-0.084	-0.084	0.949	0.945	0.872	0.972
MEAN	0.110	0.119	0.222	-0.098	-0.098	0.963	0.904	0.887	0.948	0.108	0.117	0.222	-0.091	-0.094	0.963	0.901	0.884	0.948
SD.	0.031	0.026	0.057	0.013	0.013	0.017	0.062	0.045	0.056	0.031	0.026	0.057	0.012	0.008	0.017	0.064	0.050	0.056
Min	0.054	0.060	0.165	-0.118	-0.122	0.937	0.676	0.767	0.732	0.052	0.060	0.165	-0.115	-0.110	0.937	0.675	0.745	0.732
Max	0.207	0.182	0.430	-0.078	-0.078	1.006	0.982	0.972	0.978	0.207	0.179	0.430	-0.072	-0.080	1.006	0.981	0.973	0.978

ตารางภาคผนวกที่ 10 ค่าดัชนี RMSE BIAS และสัมประสิทธิ์สหสัมพันธ์ของเงื่อนไขความสัมพันธ์ระหว่างมิติความสามารถระดับกลางในโครงสร้างที่ซับซ้อนของวิธี M และ วิธี NOP สำหรับการแปลงคะแนนจากระดับชั้นที่ 2 ไป 3 จากการทำซ้ำ 20 ครั้ง

ลำดับในการทำซ้ำ	L_23 CORR 0.5 MS																	
	MS_M									MS_NOP								
	RMSE			BIAS_TRUE			Correlation Coefficient			RMSE			BIAS_TRUE			Correlation Coefficient		
	a1	a2	d	a1	a2	d	a1	a2	d	a1	a1	d	a1	a1	d	a1	a1	d
1	0.132	0.112	0.199	-0.116	-0.116	0.481	0.690	0.782	0.966	0.128	0.103	0.199	-0.107	-0.101	0.481	0.692	0.786	0.966
2	0.124	0.127	0.170	-0.115	-0.112	0.427	0.860	0.662	0.973	0.123	0.114	0.170	-0.110	-0.094	0.427	0.860	0.637	0.973
3	0.147	0.129	0.171	-0.145	-0.101	0.434	0.873	0.817	0.974	0.130	0.129	0.171	-0.118	-0.099	0.434	0.866	0.816	0.974
4	0.127	0.099	0.185	-0.099	-0.126	0.452	0.808	0.736	0.982	0.127	0.086	0.185	-0.099	-0.109	0.452	0.808	0.734	0.982
5	0.132	0.137	0.212	-0.107	-0.105	0.437	0.837	0.733	0.986	0.125	0.137	0.212	-0.094	-0.105	0.437	0.877	0.742	0.986
6	0.112	0.094	0.181	-0.099	-0.112	0.493	0.879	0.757	0.742	0.112	0.091	0.181	-0.100	-0.105	0.493	0.986	0.877	0.742
7	0.126	0.140	0.200	-0.114	-0.105	0.470	0.873	0.766	0.974	0.126	0.140	0.200	-0.114	-0.105	0.470	0.875	0.773	0.974
8	0.134	0.140	0.178	-0.135	-0.096	0.464	0.826	0.831	0.967	0.128	0.121	0.178	-0.121	-0.075	0.464	0.830	0.819	0.967
9	0.142	0.142	0.218	-0.141	-0.116	0.458	0.821	0.745	0.961	0.134	0.134	0.218	-0.046	-0.049	0.458	0.821	0.751	0.961
10	0.149	0.122	0.172	-0.112	-0.117	0.495	0.724	0.888	0.958	0.146	0.118	0.172	-0.025	-0.054	0.495	0.710	0.890	0.958
11	0.123	0.145	0.182	-0.089	-0.138	0.461	0.785	0.694	0.962	0.116	0.143	0.182	0.002	-0.077	0.461	0.791	0.724	0.962
12	0.114	0.144	0.172	-0.091	-0.122	0.459	0.898	0.786	0.970	0.112	0.141	0.172	-0.006	-0.058	0.459	0.888	0.781	0.970
13	0.133	0.130	0.208	-0.095	-0.121	0.475	0.801	0.876	0.969	0.132	0.119	0.208	-0.091	-0.104	0.475	0.804	0.872	0.969
14	0.199	0.141	0.171	-0.154	-0.070	0.453	0.836	0.803	0.975	0.172	0.141	0.171	-0.036	-0.017	0.453	0.837	0.799	0.975
15	0.152	0.134	0.179	-0.143	-0.110	0.468	0.799	0.843	0.978	0.149	0.117	0.179	-0.053	-0.037	0.468	0.790	0.826	0.978
16	0.108	0.156	0.152	-0.082	-0.144	0.437	0.980	0.983	0.983	0.107	0.133	0.152	-0.004	-0.059	0.437	0.982	0.983	0.983
17	0.148	0.119	0.191	-0.106	-0.151	0.466	0.775	0.787	0.950	0.147	0.096	0.191	-0.030	-0.075	0.466	0.776	0.790	0.950
18	0.133	0.104	0.196	-0.106	-0.108	0.462	0.805	0.769	0.972	0.130	0.099	0.196	-0.018	-0.046	0.462	0.802	0.769	0.972
19	0.115	0.145	0.170	-0.141	-0.102	0.441	0.808	0.758	0.957	0.114	0.143	0.170	-0.056	-0.041	0.441	0.808	0.763	0.957
20	0.125	0.134	0.114	-0.105	-0.123	0.477	0.872	0.732	0.945	0.125	0.128	0.114	-0.023	-0.056	0.477	0.873	0.742	0.945
MEAN	0.134	0.130	0.181	-0.115	-0.115	0.461	0.827	0.787	0.957	0.129	0.122	0.181	-0.062	-0.073	0.461	0.834	0.794	0.957
SD.	0.020	0.017	0.023	0.021	0.018	0.019	0.063	0.073	0.052	0.015	0.018	0.023	0.043	0.028	0.019	0.073	0.074	0.052
Min	0.108	0.094	0.114	-0.154	-0.151	0.427	0.690	0.662	0.742	0.107	0.086	0.114	-0.121	-0.109	0.427	0.692	0.637	0.742
Max	0.199	0.156	0.218	-0.082	-0.070	0.495	0.980	0.983	0.986	0.172	0.143	0.218	0.002	-0.017	0.495	0.986	0.983	0.986

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางภาคผนวกที่ 11 ค่าดัชนี RMSE BIAS และสัมประสิทธิ์สหสัมพันธ์ของเงื่อนไขความสัมพันธ์ระหว่างมิติความสามารถระดับสูงในโครงสร้างที่ไม่ซับซ้อนของวิธี M และ วิธี NOP สำหรับการแปลงคะแนนจากระดับชั้นที่ 2 ไป 3 จากการทำซ้ำ 20 ครั้ง

ลำดับในการทำซ้ำ	L_23 CORR 0.8 APSS																	
	APSS_M									APSS_NOP								
	RMSE			BIAS_TRUE			Correlation Coefficient			RMSE			BIAS_TRUE			Correlation Coefficient		
	a1	a2	d	a1	a2	d	a1	a2	d	a1	a1	d	a1	a1	d	a1	a1	d
1	0.175	0.160	0.222	-0.080	-0.089	1.041	0.530	0.647	0.957	0.149	0.151	0.222	-0.057	-0.075	1.041	0.517	0.641	0.957
2	0.143	0.150	0.199	-0.061	-0.093	1.030	0.762	0.567	0.980	0.121	0.143	0.199	-0.042	-0.080	1.030	0.782	0.615	0.980
3	0.160	0.160	0.193	-0.053	-0.075	1.022	0.726	0.648	0.966	0.152	0.145	0.193	-0.038	-0.057	1.022	0.715	0.655	0.966
4	0.138	0.143	0.190	-0.089	-0.052	1.009	0.779	0.697	0.973	0.109	0.136	0.190	-0.061	-0.041	1.009	0.774	0.701	0.973
5	0.138	0.129	0.166	-0.064	-0.069	1.022	0.793	0.775	0.980	0.138	0.103	0.166	-0.060	-0.049	1.022	0.785	0.799	0.980
6	0.097	0.107	0.170	-0.096	-0.108	1.148	0.839	0.810	0.981	0.095	0.105	0.170	-0.091	-0.103	1.148	0.838	0.804	0.981
7	0.207	0.131	0.148	-0.108	-0.065	1.039	0.580	0.794	0.980	0.174	0.129	0.148	-0.068	-0.060	1.039	0.565	0.787	0.980
8	0.115	0.156	0.143	-0.048	-0.068	1.011	0.821	0.613	0.980	0.113	0.153	0.143	-0.042	-0.061	1.011	0.816	0.612	0.980
9	0.124	0.156	0.202	-0.054	-0.072	1.007	0.773	0.757	0.965	0.121	0.145	0.202	-0.048	-0.055	1.007	0.767	0.718	0.965
10	0.131	0.137	0.193	-0.051	-0.076	1.012	0.768	0.780	0.975	0.131	0.127	0.193	-0.050	-0.060	1.012	0.767	0.760	0.975
11	0.143	0.145	0.188	-0.042	-0.076	0.990	0.710	0.740	0.971	0.143	0.129	0.188	-0.039	-0.057	0.990	0.703	0.752	0.971
12	0.125	0.140	0.167	-0.046	-0.077	1.009	0.772	0.756	0.976	0.120	0.132	0.167	-0.038	-0.064	1.009	0.772	0.741	0.976
13	0.119	0.102	0.177	-0.078	-0.058	1.049	0.808	0.848	0.980	0.116	0.102	0.177	-0.070	-0.059	1.049	0.795	0.850	0.980
14	0.114	0.151	0.200	-0.064	-0.067	1.033	0.805	0.714	0.970	0.104	0.150	0.200	-0.052	-0.060	1.033	0.805	0.714	0.970
15	0.147	0.145	0.172	-0.061	-0.078	1.042	0.815	0.694	0.970	0.147	0.142	0.172	-0.061	-0.069	1.042	0.692	0.691	0.972
16	0.139	0.131	0.228	-0.060	-0.070	1.012	0.762	0.802	0.955	0.135	0.122	0.228	-0.051	-0.056	1.012	0.753	0.788	0.955
17	0.150	0.104	0.160	-0.082	-0.057	1.043	0.694	0.838	0.981	0.141	0.100	0.160	-0.066	-0.051	1.043	0.676	0.840	0.981
18	0.116	0.156	0.162	-0.047	-0.081	1.045	0.771	0.719	0.980	0.113	0.140	0.162	-0.042	-0.059	1.045	0.775	0.704	0.980
19	0.144	0.151	0.219	-0.066	-0.078	1.028	0.691	0.693	0.963	0.133	0.130	0.219	-0.051	-0.055	1.028	0.678	0.682	0.963
20	0.185	0.128	0.162	-0.088	-0.032	1.016	0.585	0.758	0.976	0.177	0.126	0.162	-0.073	-0.028	1.016	0.565	0.758	0.976
MEAN	0.141	0.139	0.183	-0.067	-0.072	1.030	0.739	0.733	0.973	0.132	0.131	0.183	-0.055	-0.060	1.030	0.727	0.731	0.973
SD.	0.026	0.018	0.024	0.019	0.016	0.032	0.086	0.075	0.008	0.022	0.017	0.024	0.014	0.015	0.032	0.089	0.070	0.008
Min	0.097	0.102	0.143	-0.108	-0.108	0.990	0.530	0.567	0.955	0.095	0.100	0.143	-0.091	-0.103	0.990	0.517	0.612	0.955
Max	0.207	0.160	0.228	-0.042	-0.032	1.148	0.839	0.848	0.981	0.177	0.153	0.228	-0.038	-0.028	1.148	0.838	0.850	0.981

ตารางภาคผนวกที่ 12 ค่าดัชนี RMSE BIAS และสัมประสิทธิ์สหสัมพันธ์ของเงื่อนไขความสัมพันธ์ระหว่างมิติความสามารถระดับสูงในโครงสร้างที่ซับซ้อนของวิธี M และ วิธี NOP สำหรับการแปลงคะแนนจากระดับชั้นที่ 2 ไป 3 จากการทำซ้ำ 20 ครั้ง

ลำดับในการทำซ้ำ	L_23 CORR 0.8 MS																	
	MS_M									MS_NOP								
	RMSE			BIAS_TRUE			Correlation Coefficient			RMSE			BIAS_TRUE			Correlation Coefficient		
	a1	a2	d	a1	a2	d	a1	a2	d	a1	a1	d	a1	a1	d	a1	a1	d
1	0.207	0.164	0.303	-0.180	-0.190	1.134	0.368	0.525	0.891	0.198	0.124	0.303	-0.166	-0.170	1.134	0.368	0.525	0.891
2	0.165	0.136	0.156	-0.160	-0.067	0.738	0.535	0.749	0.975	0.150	0.121	0.156	-0.145	-0.053	0.738	0.527	0.744	0.975
3	0.165	0.136	0.156	-0.160	-0.067	0.738	0.535	0.749	0.975	0.150	0.121	0.156	-0.145	-0.053	0.738	0.527	0.744	0.975
4	0.217	0.165	0.113	-0.152	-0.074	0.700	0.269	0.566	0.987	0.143	0.152	0.113	-0.115	-0.062	0.700	0.655	0.580	0.987
5	0.108	0.143	0.109	-0.032	-0.169	0.708	0.813	0.717	0.991	0.108	0.120	0.109	-0.031	-0.145	0.708	0.810	0.745	0.991
6	0.189	0.176	0.172	-0.119	-0.072	0.685	0.544	0.670	0.972	0.156	0.160	0.172	-0.094	-0.056	0.685	0.588	0.694	0.972
7	0.158	0.206	0.160	-0.066	-0.156	0.704	0.632	0.485	0.976	0.140	0.178	0.160	-0.054	-0.127	0.704	0.674	0.483	0.976
8	0.042	0.044	0.099	-0.098	-0.073	0.727	0.970	0.979	0.990	0.042	0.043	0.099	-0.097	-0.075	0.727	0.971	0.980	0.990
9	0.178	0.144	0.169	-0.175	-0.058	0.708	0.414	0.685	0.979	0.160	0.125	0.169	-0.156	-0.044	0.708	0.365	0.728	0.979
10	0.139	0.143	0.164	-0.101	-0.123	0.749	0.732	0.656	0.974	0.132	0.134	0.164	-0.091	-0.114	0.749	0.711	0.656	0.974
11	0.172	0.149	0.128	-0.152	-0.071	0.758	0.587	0.598	0.985	0.152	0.125	0.128	-0.131	-0.056	0.758	0.615	0.650	0.985
12	0.177	0.149	0.166	-0.175	-0.029	0.684	0.538	0.705	0.973	0.156	0.145	0.166	-0.154	-0.024	0.684	0.568	0.717	0.973
13	0.160	0.225	0.205	-0.118	-0.092	0.700	0.641	0.416	0.962	0.146	0.186	0.205	-0.103	-0.058	0.700	0.649	0.419	0.962
14	0.117	0.119	0.135	-0.076	-0.127	0.708	0.802	0.782	0.983	0.105	0.116	0.135	-0.065	-0.120	0.708	0.831	0.760	0.983
15	0.200	0.163	0.195	-0.130	-0.094	0.700	0.457	0.694	0.963	0.173	0.137	0.195	-0.109	-0.071	0.700	0.509	0.728	0.963
16	0.197	0.163	0.148	-0.192	-0.040	0.727	0.425	0.660	0.982	0.159	0.143	0.148	-0.159	-0.026	0.727	0.431	0.721	0.982
17	0.192	0.172	0.135	-0.125	-0.112	0.701	0.622	0.600	0.985	0.136	0.156	0.135	-0.083	-0.096	0.701	0.686	0.618	0.985
18	0.203	0.208	0.176	-0.118	-0.150	0.746	0.508	0.425	0.971	0.138	0.165	0.176	-0.078	-0.113	0.746	0.702	0.424	0.971
19	0.174	0.154	0.156	-0.115	-0.056	0.718	0.547	0.674	0.978	0.146	0.146	0.156	-0.095	-0.046	0.718	0.625	0.689	0.978
20	0.189	0.160	0.134	-0.150	-0.052	0.715	0.459	0.622	0.983	0.159	0.142	0.134	-0.124	-0.040	0.715	0.448	0.702	0.983
MEAN	0.167	0.156	0.159	-0.130	-0.094	0.737	0.570	0.648	0.974	0.142	0.137	0.159	-0.110	-0.077	0.737	0.613	0.665	0.974
SD.	0.041	0.037	0.043	0.041	0.045	0.096	0.166	0.129	0.021	0.031	0.030	0.043	0.038	0.041	0.096	0.155	0.131	0.021
Min	0.042	0.044	0.099	-0.192	-0.190	0.684	0.269	0.416	0.891	0.042	0.043	0.099	-0.166	-0.170	0.684	0.365	0.419	0.891
Max	0.217	0.225	0.303	-0.032	-0.029	1.134	0.970	0.979	0.991	0.198	0.186	0.303	-0.031	-0.024	1.134	0.971	0.980	0.991

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวพัชรี จันทร์เพ็ง เกิดเมื่อวันที่ 9 พฤษภาคม พ.ศ. 2523 ที่อำเภอเมือง จังหวัดสกลนคร สำเร็จการศึกษาปริญญาศึกษาศาสตรบัณฑิต (เกียรตินิยม) สาขาการมัธยมศึกษา วิชาเอกคณิตศาสตร์ศึกษา จากมหาวิทยาลัยขอนแก่น ในปีการศึกษา 2545 โดยได้รับทุนในโครงการเร่งรัดการผลิตและพัฒนาบัณฑิต ระดับปริญญาตรี สาขาคณิตศาสตร์ของประเทศ (รพค.) ต่อมาได้รับทุนโครงการผลิตและพัฒนาอาจารย์ (UDC.) จากภาควิชาประเมินผลและวิจัยการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น เพื่อเข้าศึกษาต่อในหลักสูตรครุศาสตรมหาบัณฑิตและหลักสูตรครุศาสตรดุษฎีบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2546 และปีการศึกษา 2548 ตามลำดับ ปัจจุบันเป็นอาจารย์ประจำสาขาวิชาการวัดและประเมินผลการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย