การพัฒนาเกณฑ์ตัดสินคุณภาพการปรับเทียบคะแนน ตามทฤษฎีการตอบสนองข้อสอบ

นายอดิศร ศรีบุญวงษ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาครุศาสตรดุษฎีบัณฑิต
สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยการศึกษา
คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2545

ISBN 974-17-1634-6

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

I20877808

THE DEVELOPMENT OF CRITERIA TO EVALUATE THE QUALITY OF TEST EQUATING BASED ON ITEM RESPONSE TEST THEORY

Mr. Adisorn Sriboonwong

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in Educational Measurement and Evaluation
Department of Educational Research
Faculty of Education
Chulalongkorn University
Academic Year 2002

ISBN 974-17-1634-6

หัวข้อวิทยานิพนธ์	การพัฒนาเกณฑ์ตัดสินคุณภาพการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ
โดย	นายอดิศร ศรีบุญวงษ์
สาขาวิชา	การวัดและประเมินผลการศึกษา
อาจารย์ที่ปรึกษา	รองศาสตราจารย์ ดร.ศิริเดช สุชีวะ
อาจารย์ที่ปรึกษาร่วม	รองศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาดุษฎีบัณฑิต

.....คณบดีคณะครุศาสตร์
(รองศาสตราจารย์ ดร.ไพฑูรย์ สินลารัตน์)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ
(รองศาสตราจารย์ ดร.ทวีวัฒน์ ปิตยานนท์)

.....อาจารย์ที่ปรึกษา
(รองศาสตราจารย์ ดร.ศิริเดช สุชีวะ)

.....อาจารย์ที่ปรึกษาร่วม
(รองศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี)

.....กรรมการ
(อาจารย์ ดร.ชูศักดิ์ ขัมภลิขิต)

.....กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.สุชาดา บวรกิติวงศ์)

อดิศร ศรีบุญวงษ์: การพัฒนาเกณฑ์ตัดสินคุณภาพการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ (THE DEVELCPMENT OF CRITERIA TO EVALUATE THE QUALITY OF TEST EQUATING BASED ON ITEM RESPONSE TEST THEORY) อ.ที่ปรึกษา: รศ. ดร.ศิริเดช สุชีวะ, อ.ที่ปรึกษาร่วม รศ. ดร.ศิริชัย กาญจนวาสี 155 หน้า. ISBN 974-17-1634-6

การวิจัยครั้งนี้มีวัตถุประสงค์ (1) เพื่อพัฒนาเกณฑ์ตัดสินคุณภาพการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ (2) เพื่อตรวจสอบคุณภาพของเกณฑ์ที่พัฒนาขึ้น ด้วยการหาความสอดคล้องของผลการตัดสินคุณภาพการปรับเทียบคะแนนด้วยเกณฑ์ที่พัฒนาขึ้นกับเกณฑ์ความเสมอภาคของลอร์ด และเกณฑ์ของปิเตอร์เซนและคณะ โดยใช้ข้อมูลจำลองจากโปรแกรม IRTDATA ที่ตรวจสอบความถูกต้องโดยการวิเคราะห์ซ้ำด้วยโปรแกรม BILOG การจำลองข้อมูลตามเงื่อนไขการปรับเทียบคะแนนด้วย โมเดลการตอบสนองข้อสอบแบบ 1 และ 3 พารามิเตอร์ วิเคราะห์คุณภาพของการปรับเทียบคะแนนด้วยการปรับเทียบกลับสู่แบบสอบเดิมและการใช้กลุ่มสอบทานผล เก็บรวบรวมข้อมูลโดยใช้กลุ่มสมมูลและการใช้ข้อสอบร่วม วิธีปรับเทียบคะแนน 4 วิธี คือ วิธี The Same Scaling Convention, วิธี Mean and Mean, วิธี Mean and Sigma, และ วิธี Characteristic Curve ที่วิเคราะห์ด้วยโปรแกรม EQUATE ผู้สอบจำนวน 100, 200 300,...,3,000 คน และข้อสอบจำนวน 10, 20, 30,..., 200 ข้อ จัดกลุ่มค่าดัชนีความแตกต่าง AMD, MAD, และ RMS โดยใช้จุดตัดซึ่งเป็นค่าดัชนีความแตกต่างเมื่อทดสอบความแตกต่างระหว่างค่าเฉลี่ยของคะแนนจริงที่ปรับเทียบแล้วกับคะแนนจริงที่ยังไม่ได้ปรับเทียบ สำหรับผู้สอบกลุ่มเดียวกันด้วยสถิติ t-test แล้วพบค่าวิกฤต ที่ระดับนัยสำคัญทางสถิติที่ระดับ .001, .01 และ .02 คำนวณค่าดัชนีความแตกต่างแล้วนำกลุ่มค่าดัชนีความแตกต่างแต่ละกลุ่มมาระบุระดับคุณภาพการปรับเทียบคะแนนเพื่อกำหนดเป็นเกณฑ์

ผลสรุปของการวิจัยที่สำคัญมีดังนี้

1. ได้เกณฑ์ตัดสินคุณภาพการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบที่พัฒนาขึ้น มีทั้งหมด 7 เกณฑ์ คือ เกณฑ์รวมทุกเงื่อนไข เกณฑ์สำหรับโมเดล 1 พารามิเตอร์ เกณฑ์สำหรับโมเดล 3 พารามิเตอร์ เกณฑ์สำหรับการใช้กลุ่มสมมูล เกณฑ์สำหรับการใช้แบบสอบร่วม เกณฑ์สำหรับการปรับเทียบกลับสู่แบบสอบเดิม และเกณฑ์สำหรับการใช้กลุ่มสอบทานผล แต่ละเกณฑ์เป็นกลุ่มค่าดัชนี AMD, MAD, และ RMS ที่บอกระดับคุณภาพการปรับเทียบคะแนน 4 ระดับ คือ การปรับเทียบน่าพอใจอย่างยิ่ง การปรับเทียบน่าพอใจ การปรับเทียบไม่น่าพอใจ และการปรับเทียบไม่น่าพอใจอย่างยิ่ง

2. ผลการตัดสินคุณภาพการปรับเทียบคะแนนเมื่อใช้เกณฑ์ที่พัฒนาขึ้นเปรียบเทียบกับผลการใช้ตามเกณฑ์ของปิเตอร์เซนและคณะพบว่าไม่สอดคล้องกัน แต่เมื่อเทียบกับเกณฑ์ความเสมอภาคของลอร์ดปรากฏว่าผลการตัดสินคุณภาพการปรับเทียบมีความสอดคล้องกัน

ภาควิชา **วิจัยการศึกษา** ลายมือชื่อนิสิต.....

สาขาวิชา **การวัดและประเมินผลการศึกษา** ลายมือชื่ออาจารย์ที่ปรึกษา.....

ปีการศึกษา **2545** ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....

4184957727: MAJOR EDUCATIONAL MEASUREMENT AND EVALUATION
KEY WORD: TEST EQUATING BASED ON ITEM RESPONSE THEORY/ DESIGN FOR EQUATING/ COMMON ITEM/ INDICES/ EQUATING A TEST TO ITSELF/ CROSS-VALIDATION SAMPLES

ADISORN SRIBOONWONG: THE DEVELOPMENT OF CRITERIA TO EVALUATE THE QUALITY OF TEST EQUATING BASED ON ITEM RESPONSE TEST THEORY. DISSERTATION ADVISOR: ASSOC. PROF. SIRIDEJ SUJIVA, Ph.D., DISSERTATION CO-ADVISOR: ASSOC. PROF. SIRICHAI KANJANAWASEE, Ph.D. 155 pp.
ISBN 974-17-1634-6

In order to equate tests under Item Response Theory (IRT), the results need to be evaluated and equating criteria has been used to decide the quality of test equating. The purpose of this study was to develop criteria from simulated data for three statistical indices: absolute mean difference (AMD), mean absolute difference (MAD), and root mean square (RMS), using the item response data simulated by IRTDATA program. The criteria identifying the quality of test equating for each index was developed by respective critical values of hypothesis's t-test from each condition of test equating: one-parameter and three-parameter logistic model, equivalent groups equating design and common-item nonequivalent groups design; four test equating methods: The Same Scaling Convention Method, Mean and Mean Method, Mean and Sigma, Method, and Characteristic Curve Method; equating a test to itself and cross validation samples; 20 test lengths: 10, 20, ... , 200 items, and 30 sizes of examinees: 100, 200,... , 3,000.

Major findings were as follow:

1. Seven sets of criteria of each index were set based on condition of test equating: criteria for overall conditions of test equating, criteria for one-parameter logistic model, criteria for three-parameter logistic model, criteria for equivalent groups equating design, criteria for common-item nonequivalent groups equating design, criteria for equating a test to itself, and criteria for employing cross-validation samples. The three indices: AMD, MAD, and RMS that they indicated four level of test equating quality: very satisfactory equating, satisfactory equating, unsatisfactory equating, and very unsatisfactory equating.

2. There were inconsistency for quality of test equating from both these criterion and Petersen's criteria whereas there were consistency for quality of test equating with Lord's equating criteria.

Department Educational research Student's signature Adisorn S.
Field of study Educational measurement and evaluation Advisor's signature
Academic year 2002 Co-advisor's signature Sirichai Kanjanawasee

กิตติกรรมประกาศ

การทำวิทยานิพนธ์ฉบับนี้ ผู้วิจัยได้รับความรู้และคำแนะนำต่าง ๆ ด้วยความกรุณาอย่างยิ่งจากอาจารย์ที่ปรึกษาทั้งสองท่าน คือ รศ.ดร.ศิริเดช สุชีวะ และ รศ.ดร.ศิริชัย กาญจนวาสี และได้รับคำแนะนำในการดำเนินการวิจัยเพิ่มเติม ทำให้วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงด้วยดีจาก รศ.ดร.ทวีวัฒน์ ปิตยานนท์ อาจารย์ ดร.ชูศักดิ์ ขัมภลิขิต และ ผศ.ดร.สุชาดา บวรกิติวงศ์ ผู้วิจัยขอขอบพระคุณอย่างสูงไว้ ณ ที่นี้

ขอขอบพระคุณ Mr. George Johanson จาก Ohio University Athens ที่อนุเคราะห์โปรแกรม IRTDATA และให้คำแนะนำการใช้โปรแกรมที่เป็นประโยชน์กับการวิจัยในครั้งนี้

ขอขอบพระคุณ ผศ.ดร.สุภมาส อังศุโชติ จากมหาวิทยาลัยสุโขทัยธรรมาธิราช รอ.พงศ์เทพ จิระโร ที่คอยช่วยเหลือและให้กำลังใจในการทำวิทยานิพนธ์นี้

ขอขอบพระคุณจุฬาลงกรณ์มหาวิทยาลัยที่ให้ทุนอุดหนุนการทำวิทยานิพนธ์ฉบับนี้

ขอขอบพระคุณครู-อาจารย์ ที่ประสิทธิ์ประสาทความรู้ ให้แนวคิดและแนวทางในการแสวงหาความรู้ตั้งแต่อดีตจนถึงปัจจุบัน

ท้ายที่สุดนี้ขอขอบพระคุณ บิดา มารดา ที่เป็นกำลังใจในการศึกษา สนับสนุนและให้ความช่วยเหลือด้วยดีเสมอมา

สารบัญ

หน้า

บทคัดย่อภาษาไทย..... ง
บทคัดย่อภาษาอังกฤษ..... จ
กิตติกรรมประกาศ..... ฉ
สารบัญ..... ช
สารบัญตาราง..... ฌ
สารบัญภาพ..... ฒ
บทที่
1 บทนำ..... 1
ความเป็นมาและความสำคัญของปัญหา..... 1
วัตถุประสงค์ของการวิจัย.....11
ขอบเขตของการวิจัย.....11
คำนิยามเชิงปฏิบัติการที่ใช้ในการวิจัย.....12
ประโยชน์ที่คาดว่าจะได้รับ.....14

2 เอกสารและงานวิจัยที่เกี่ยวข้อง.....15
ตอนที่ 1 แนวคิดเชิงทฤษฎีของการปรับเทียบคะแนนระหว่างแบบสอบ.....16
ตอนที่ 2 ประเภทของการปรับเทียบคะแนน.....17
ตอนที่ 3 ขั้นตอนการดำเนินการปรับเทียบคะแนน.....18
ตอนที่ 4 การออกแบบวิธีการเก็บรวบรวมข้อมูล.....21
ตอนที่ 5 วิธีการปรับเทียบคะแนนตามทฤษฎีการวัดแบบดั้งเดิม.....22
ตอนที่ 6 การปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ.....25
ตอนที่ 7 เกณฑ์ที่ใช้ในการปรับเทียบคะแนน.....40
ตอนที่ 8 งานวิจัยที่เกี่ยวข้องกับการปรับเทียบคะแนน.....49

3 วิธีดำเนินการวิจัย.....55
ตอนที่ 1 ตรวจสอบคุณภาพของการจำลองข้อมูลจากโปรแกรม IRTDATA.....55
ตอนที่ 2 จำลองข้อมูลเพื่อดำเนินการปรับเทียบคะแนนตามแบบแผนที่กำหนด
จัดกลุ่มดัชนีความแตกต่างและวิเคราะห์หาค่าดัชนีความแตกต่าง.....57
ตอนที่ 3 พัฒนาเกณฑ์จากดัชนีความแตกต่าง.....70
ตอนที่ 4 ตรวจสอบคุณภาพของเกณฑ์ที่พัฒนาขึ้น.....72

สารบัญ (ต่อ)

บทที่ หน้า

4 ผลการวิเคราะห์ข้อมูล....74

ตอนที่ 1 ผลการตรวจสอบความถูกต้องและความเป็นเอกมิติของข้อมูลที่จำลองจากโปรแกรม IRTDATA....74

ตอนที่ 2 ผลการวิเคราะห์จุดตัดแบ่งกลุ่มดัชนีความแตกต่างออกเป็นกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงและกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำ....81

ตอนที่ 3 เกณฑ์ตัดสินคุณภาพการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบที่พัฒนาขึ้น....86

ตอนที่ 4 ผลการตรวจสอบคุณภาพของเกณฑ์ที่พัฒนาขึ้น....107

5 สรุปผลการวิจัย อภิปรายผลและข้อเสนอแนะ....117

สรุปผลการวิจัย....118

อภิปรายผลการวิจัย....123

ข้อเสนอแนะในการนำเกณฑ์ที่พัฒนาขึ้นไปใช้....128

ข้อเสนอแนะในการวิจัยครั้งต่อไป....129

รายการอ้างอิง....130

ภาคผนวก....135

ภาคผนวก ก การใช้โปรแกรม IRTDATA....136

ภาคผนวก ข การใช้โปรแกรม BILOG....144

ภาคผนวก ค การใช้โปรแกรม EQUATE....147

ภาคผนวก ง โค้งคุณลักษณะข้อสอบ....151

ประวัติผู้วิจัย....155

สารบัญตาราง

ตารางที่ หน้า

1 พารามิเตอร์ข้อสอบและค่าความสามารถของผู้สอบสำหรับ 2 สเกลของแบบสอบ.....32
2 ความแตกต่างระหว่างค่าเฉลี่ยพารามิเตอร์ผู้สอบ 300 คน พารามิเตอร์ข้อสอบ 40 ข้อ จากโปรแกรม IRTDATA และโปรแกรม BILOG สำหรับโมเดล 1 พารามิเตอร์.....75
3. ความแตกต่างระหว่างค่าเฉลี่ยพารามิเตอร์ผู้สอบ 1,000 คน พารามิเตอร์ข้อสอบ 80 ข้อ จากโปรแกรม IRTDATA และโปรแกรม BILOG สำหรับโมเดล 1 พารามิเตอร์....75
4. ความแตกต่างระหว่างค่าเฉลี่ยพารามิเตอร์ผู้สอบ 2,000 คน พารามิเตอร์ข้อสอบ 40 ข้อ จากโปรแกรม IRTDATA และโปรแกรม BILOG สำหรับโมเดล 3 พารามิเตอร์.....76
5. ความแตกต่างระหว่างค่าเฉลี่ยพารามิเตอร์ผู้สอบ 500 คน พารามิเตอร์ข้อสอบ 60 ข้อ จากโปรแกรม IRTDATA และโปรแกรม BILOG สำหรับโมเดล 3 พารามิเตอร์.....77
6. ค่าไอเกนและร้อยละของความแปรปรวนของตัวประกอบ สำหรับข้อมูลที่ได้จากโปรแกรม IRTDATA สำหรับโมเดล 1 พารามิเตอร์ ข้อสอบ 40 ข้อ ผู้สอบ 300 คน.....78
7 ค่าไอเกนและร้อยละของความแปรปรวนของตัวประกอบ สำหรับข้อมูลที่ได้จากโปรแกรม IRTDATA สำหรับโมเดล 1 พารามิเตอร์ ข้อสอบ 80 ข้อ ผู้สอบ 1,000 คน...78
8 ค่าไอเกนและร้อยละของความแปรปรวนของตัวประกอบ สำหรับข้อมูลที่ได้จากโปรแกรม IRTDATA สำหรับโมเดล 3 พารามิเตอร์ ข้อสอบ 40 ข้อ ผู้ตอบ 2,000 คน..79
9 ค่าไอเกนและร้อยละของความแปรปรวนของตัวประกอบ สำหรับข้อมูลที่ได้จากโปรแกรม IRTDATA สำหรับโมเดล 3 พารามิเตอร์ ข้อสอบ 60 ข้อ ผู้ตอบ 500 คน.....80
10 การทดสอบความแตกต่างของค่าเฉลี่ยคะแนนจริงที่ไม่ได้ปรับและคะแนนจริงที่ปรับเทียบแล้ว สำหรับผู้สอบกลุ่มเดียวกัน จำนวน 1,200 คน แบบสอบจำนวน 60 ข้อ โมเดล 3 พารามิเตอร์ ปรับเทียบคะแนนด้วยวิธี Mean and Sigma เก็บรวบรวมข้อมูลด้วยการใช้ข้อสอบร่วม หาคุณภาพการปรับเทียบโดยการใช้กลุ่มสอบทานผล.....81
11 ดัชนีความแตกต่าง AMD, MAD และ RMS ที่กำหนดเป็นจุดตัดแบ่งกลุ่มดัชนีออกเป็นกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงและกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำ ใช้ผลจากตารางที่ 10 82
12 การทดสอบความแตกต่างของค่าเฉลี่ยคะแนนจริงที่ไม่ได้ปรับและคะแนนจริงที่ปรับเทียบแล้ว สำหรับผู้สอบกลุ่มเดียวกัน จำนวน 900 คน แบบสอบจำนวน 40 ข้อ โมเดล 1 พารามิเตอร์ ปรับเทียบคะแนนด้วยวิธี Mean and Sigma เก็บรวบรวมข้อมูลด้วยการใช้ข้อสอบร่วม หาคุณภาพการปรับเทียบโดยการปรับเทียบกลับสู่แบบสอบเดิม.....82

สารบัญตาราง (ต่อ)

ตารางที่ หน้า

13 ดัชนีความแตกต่าง AMD, MAD และ RMS ที่กำหนดเป็นจุดตัดแบ่งกลุ่มดัชนีออกเป็นกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงและกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำ ใช้ผลจากตารางที่ 12 83

14 การทดสอบความแตกต่างของค่าเฉลี่ยคะแนนจริงที่ไม่ได้ปรับและคะแนนจริงที่ปรับเทียบแล้ว สำหรับผู้สอบกลุ่มเดียวกัน จำนวน 1,000 คน แบบสอบจำนวน 30 ข้อ โมเดล 3 พารามิเตอร์ ปรับเทียบคะแนนด้วยวิธีการกำหนดค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ผู้สอบให้เท่ากันก่อนการวิเคราะห์ เก็บรวบรวมข้อมูลด้วยการใช้กลุ่มสมมูล หาคุณภาพการปรับเทียบโดยการใช้กลุ่มสอบทานผล.....83

15 ดัชนีความแตกต่าง AMD, MAD และ RMS ที่กำหนดเป็นจุดตัดแบ่งกลุ่มดัชนีออกเป็นกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงและกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำ ใช้ผลจากตารางที่ 14 84

16 ดัชนีความแตกต่าง AMD, MAD และ RMS ที่กำหนดเป็นจุดตัดแบ่งกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงและกลุ่มที่แสดงคุณภาพการปรับเทียบระดับต่ำ จำแนกตามเงื่อนไขการปรับเทียบคะแนน.....85

17 การทดสอบความแตกต่างของค่าเฉลี่ยคะแนนจริงที่ไม่ได้ปรับและคะแนนจริงที่ปรับเทียบแล้ว สำหรับผู้สอบกลุ่มเดียวกัน จำนวน 1,400 คน แบบสอบจำนวน 80 ข้อ โมเดล 1 พารามิเตอร์ ปรับเทียบคะแนนด้วยวิธี Mean and Sigma เก็บรวบรวมข้อมูลด้วยการใช้ข้อสอบร่วม หาคุณภาพการปรับเทียบโดยการใช้กลุ่มสอบทานผล.....87

18 ดัชนีความแตกต่าง AMD, MAD และ RMS ที่กำหนดเป็นจุดตัดแบ่งกลุ่มดัชนีออกเป็นกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงและกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำ ใช้ผลจากตารางที่ 17 88

19 การทดสอบความแตกต่างของค่าเฉลี่ยคะแนนจริงที่ไม่ได้ปรับและคะแนนจริงที่ปรับเทียบแล้ว สำหรับผู้สอบกลุ่มเดียวกัน จำนวน 1,600 คน แบบสอบจำนวน 90 ข้อ โมเดล 3 พารามิเตอร์ ปรับเทียบคะแนนด้วยวิธี Mean and Sigma เก็บรวบรวมข้อมูลด้วยการใช้ข้อสอบร่วม หาคุณภาพการปรับเทียบโดยการปรับเทียบกลับสู่แบบสอบเดิม.....88

สารบัญตาราง (ต่อ)

ตารางที่ หน้า

20 ดัชนีความแตกต่าง AMD, MAD และ RMS ที่กำหนดเป็นจุดตัดแบ่งกลุ่มดัชนีออกเป็นกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงและกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำ ใช้ผลจากตารางที่ 19. 89

21 การทดสอบความแตกต่างของค่าเฉลี่ยคะแนนจริงที่ไม่ได้ปรับและคะแนนจริงที่ปรับเทียบแล้ว สำหรับผู้สอบกลุ่มเดียวกัน จำนวน 1,200 คน แบบสอบจำนวน 70 ข้อ โมเดล 3 พารามิเตอร์ ปรับเทียบคะแนนด้วยวิธีการกำหนดค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ผู้สอบให้เท่ากันก่อนการวิเคราะห์ เก็บรวบรวมข้อมูลด้วยการใช้กลุ่มสมมูล หาคุณภาพการปรับเทียบโดยการใช้กลุ่มสอบทานผล.....89

22 ดัชนีความแตกต่าง AMD, MAD และ RMS ที่กำหนดเป็นจุดตัดแบ่งกลุ่มดัชนีออกเป็นกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงและกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำ ใช้ผลจากตารางที่ 21.....90

23 การทดสอบความแตกต่างของค่าเฉลี่ยคะแนนจริงที่ไม่ได้ปรับและคะแนนจริงที่ปรับเทียบแล้ว สำหรับผู้สอบกลุ่มเดียวกัน จำนวน 1,100 คน แบบสอบจำนวน 50 ข้อ โมเดล 3 พารามิเตอร์ ปรับเทียบคะแนนด้วยวิธี Mean and Mean เก็บรวบรวมข้อมูลด้วยการใช้ข้อสอบร่วม หาคุณภาพการปรับเทียบโดยการใช้กลุ่มสอบทานผล.....91

24 ดัชนีความแตกต่าง AMD, MAD และ RMS ที่กำหนดเป็นจุดตัดแบ่งกลุ่มดัชนีออกเป็นกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงและกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำ ใช้ผลจากตารางที่ 23 91

25 การทดสอบความแตกต่างของค่าเฉลี่ยคะแนนจริงที่ไม่ได้ปรับและคะแนนจริงที่ปรับเทียบแล้ว สำหรับผู้สอบกลุ่มเดียวกัน จำนวน 1,000 คน แบบสอบจำนวน 60 ข้อ โมเดล 1 พารามิเตอร์ ปรับเทียบคะแนนด้วยวิธี Mean and Mean เก็บรวบรวมข้อมูลด้วยการใช้ข้อสอบร่วม หาคุณภาพการปรับเทียบโดยการการปรับเทียบกลับสู่แบบสอบเดิม.....92

26 ดัชนีความแตกต่าง AMD, MAD และ RMS ที่กำหนดเป็นจุดตัดแบ่งกลุ่มดัชนีออกเป็นกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงและกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำ ใช้ผลจากตารางที่ 25 92

สารบัญตาราง (ต่อ)

ตารางที่ หน้า

27 การทดสอบความแตกต่างของค่าเฉลี่ยคะแนนจริงที่ไม่ได้ปรับและคะแนนจริงที่ปรับเทียบแล้ว สำหรับผู้สอบกลุ่มเดียวกัน จำนวน 700 คน แบบสอบจำนวน 40 ข้อ โมเดล 1 พารามิเตอร์ ปรับเทียบคะแนนด้วยวิธีการกำหนดค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ผู้สอบให้เท่ากันก่อนการวิเคราะห์ เก็บรวบรวมข้อมูลด้วยการใช้กลุ่มสมมูล หาคุณภาพการปรับเทียบโดยการใช้กลุ่มสอบทานผล.....93

28 ดัชนีความแตกต่าง AMD, MAD และ RMS ที่กำหนดเป็นจุดตัดแบ่งกลุ่มดัชนีออกเป็นกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงและกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำ ใช้ผลจากตารางที่ 27.....94

29 ดัชนีความแตกต่าง AMD, MAD และ RMS ที่กำหนดเป็นจุดตัดแบ่งกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงออกเป็นกลุ่มย่อย จำแนกตามเงื่อนไขการปรับเทียบคะแนน.....94

30 ดัชนีความแตกต่าง AMD, MAD และ RMS ที่กำหนดเป็นจุดตัดแบ่งกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำออกเป็นกลุ่มย่อย จำแนกตามเงื่อนไขการปรับเทียบคะแนน.....95

31 จุดตัดที่เป็นค่าดัชนีความแตกต่างระหว่างคะแนนที่ปรับเทียบแล้วและคะแนนที่ยังไม่ได้ปรับเทียบ สำหรับเกณฑ์รวมทุกเงื่อนไข.....96

32 เกณฑ์ตัดสินคุณภาพการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบสำหรับเกณฑ์รวมทุกเงื่อนไข.....98

33 จุดตัดที่เป็นค่าดัชนีความแตกต่างระหว่างคะแนนที่ปรับเทียบแล้วและคะแนนที่ยังไม่ได้ปรับเทียบ เกณฑ์สำหรับโมเดล 1 พารามิเตอร์.....98

34 เกณฑ์ตัดสินคุณภาพการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบเกณฑ์สำหรับโมเดล 1 พารามิเตอร์.....99

35 จุดตัดที่เป็นค่าดัชนีความแตกต่างระหว่างคะแนนที่ปรับเทียบแล้วและคะแนนที่ยังไม่ได้ปรับเทียบ เกณฑ์สำหรับโมเดล 3 พารามิเตอร์.....100

36 เกณฑ์ตัดสินคุณภาพการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบเกณฑ์สำหรับโมเดล 3 พารามิเตอร์.....101

37 จุดตัดที่เป็นค่าดัชนีความแตกต่างระหว่างคะแนนที่ปรับเทียบแล้วและคะแนนที่ยังไม่ได้ปรับเทียบ เกณฑ์สำหรับแบบแผนการเก็บรวบรวมข้อมูลใช้กลุ่มสมมูล....101

สารบัญตาราง (ต่อ)

ตารางที่		หน้า
38	เกณฑ์ตัดสินคุณภาพการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ เกณฑ์สำหรับแบบแผนการเก็บรวบรวมข้อมูลใช้กลุ่มสมมูล	102
39	จุดตัดที่เป็นค่าดัชนีความแตกต่างระหว่างคะแนนที่ปรับเทียบแล้วและคะแนนที่ยังไม่ได้ปรับเทียบ เกณฑ์สำหรับแบบแผนการเก็บรวบรวมข้อมูลใช้ข้อสอบร่วม	103
40	เกณฑ์ตัดสินคุณภาพการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ เกณฑ์สำหรับแบบแผนการเก็บรวบรวมข้อมูลใช้ข้อสอบร่วม	104
41	จุดตัดที่เป็นค่าดัชนีความแตกต่างระหว่างคะแนนที่ปรับเทียบแล้วและคะแนนที่ยังไม่ได้ปรับเทียบ เกณฑ์สำหรับการหาคุณภาพการปรับเทียบด้วยการปรับเทียบกลับสู่แบบสอบเดิม	104
42	เกณฑ์ตัดสินคุณภาพการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ เกณฑ์สำหรับแบบแผนการปรับเทียบกลับสู่แบบสอบเดิม	105
43	จุดตัดที่เป็นค่าดัชนีความแตกต่างระหว่างคะแนนที่ปรับเทียบแล้วและคะแนนที่ยังไม่ได้ปรับเทียบ เกณฑ์สำหรับการหาคุณภาพการปรับเทียบด้วยการใช้กลุ่มสอบทานผล	106
44	เกณฑ์ตัดสินคุณภาพการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ เกณฑ์สำหรับการหาคุณภาพการปรับเทียบด้วยการใช้กลุ่มสอบทานผล	107
45	การทดสอบความแตกต่างของค่าเฉลี่ยคะแนนจริงที่ยังไม่ได้ปรับเทียบกับค่าเฉลี่ยคะแนนจริงที่ปรับเทียบแล้ว จำแนกตามระดับคุณภาพการปรับเทียบ จากข้อมูลที่จำลองขึ้นตามเงื่อนไขการปรับเทียบคะแนน โดยใช้เกณฑ์ที่พัฒนาขึ้น	108
46	ค่าดัชนีความแตกต่าง AMD, MAD และ RMS จำแนกตามระดับคุณภาพการปรับเทียบคะแนนและเงื่อนไขการปรับเทียบคะแนน	108
47	ค่าดัชนีความแตกต่างตามเกณฑ์ของปิเตอร์เซนและคณะ	109
48	ทดสอบความแตกต่างระหว่างการแจกแจงคะแนนจริงที่ปรับเทียบแล้วกับการการแจกแจงคะแนนจริงที่ปรับเทียบแล้ว ด้วยสถิติ Wilcoxon sign-rank test สำหรับระดับคุณภาพการปรับเทียบทั้ง 4 ระดับ	110
49	ความสอดคล้องของเกณฑ์ที่พัฒนาขึ้น สำหรับเกณฑ์รวมทุกเงื่อนไข กับเกณฑ์ความเสมอภาคของลอร์ดและเกณฑ์ของปิเตอร์เซนและคณะ	111
50	ความสอดคล้องของเกณฑ์ที่พัฒนาขึ้น เกณฑ์สำหรับโมเดล 1 พารามิเตอร์ กับเกณฑ์ความเสมอภาคของลอร์ดและเกณฑ์ของปิเตอร์เซนและคณะ	112

สารบัญตาราง (ต่อ)

ตารางที่		หน้า
51	ความสอดคล้องของเกณฑ์ที่พัฒนาขึ้น เกณฑ์สำหรับโมเดล 3 พารามิเตอร์ กับเกณฑ์ความเสมอภาคของลอร์ดและเกณฑ์ของปิเตอร์เซนและคณะ	112
52	ความสอดคล้องของเกณฑ์ที่พัฒนาขึ้น เกณฑ์สำหรับแบบแผนการเก็บรวบรวมข้อมูลด้วยการใช้ข้อสอบร่วม กับเกณฑ์ความเสมอภาคของลอร์ดและเกณฑ์ของปิเตอร์เซนและคณะ	113
53	ความสอดคล้องของเกณฑ์ที่พัฒนาขึ้น เกณฑ์สำหรับแบบแผนการเก็บรวบรวมข้อมูลด้วยการใช้กลุ่มสมมูล กับเกณฑ์ความเสมอภาคของลอร์ดและเกณฑ์ของปิเตอร์เซนและคณะ	114
54	ความสอดคล้องของเกณฑ์ที่พัฒนาขึ้น เกณฑ์สำหรับการหาคุณภาพการปรับเทียบด้วยการปรับเทียบกลับสู่แบบสอบเดิม กับเกณฑ์ความเสมอภาคของลอร์ดและเกณฑ์ของปิเตอร์เซนและคณะ	114
55	ความสอดคล้องของเกณฑ์ที่พัฒนาขึ้น เกณฑ์สำหรับการหาคุณภาพการปรับเทียบด้วยการใช้กลุ่มสอบทานผล กับเกณฑ์ความเสมอภาคของลอร์ดและเกณฑ์ของปิเตอร์เซนและคณะ	115

สารบัญภาพ

ภาพที่		หน้า
1	แผนผังการจำลองข้อมูลแบบแผนกลุ่มสมมูล	58
2	แผนผังการจำลองข้อมูลตามแบบแผนกลุ่มไม่เท่าเทียมกันใช้แบบสอบร่วม	59
3	แผนผังการจำลองข้อมูล ปรับเทียบคะแนนและคำนวณค่าดัชนีความแตกต่างตามแบบแผนกลุ่มสมมูล ใช้กลุ่มสอบทานผล	62
4	แผนผังการจำลองข้อมูล ปรับเทียบคะแนนและคำนวณค่าดัชนีความแตกต่างตามแบบแผนกลุ่มไม่เท่าเทียมกันใช้แบบสอบร่วม และใช้การปรับเทียบกลับสู่แบบสอบเดิม	66
5	แผนผังการจำลองข้อมูล ปรับเทียบคะแนนและคำนวณค่าดัชนีความแตกต่างตามแบบแผนกลุ่มไม่เท่าเทียมกันใช้แบบสอบร่วม และใช้กลุ่มสอบทานผล	69
6	โค้งคุณลักษณะข้อสอบ (ICC) ตามโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ของข้อสอบ 5 ข้อ จากแบบสอบฉบับที่ 1 ที่ไม่ได้ปรับเทียบ	151
7	โค้งคุณลักษณะข้อสอบ (ICC) ตามโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ของข้อสอบ 5 ข้อ จากแบบสอบฉบับที่ 2 ที่ปรับเทียบเข้าสู่สเกลของแบบสอบฉบับที่ 1 ด้วยวิธี Mean and Mean	151
8	โค้งคุณลักษณะข้อสอบ (ICC) ตามโมเดลการตอบสนองข้อสอบ 3 พารามิเตอร์ของข้อสอบ 5 ข้อ จากแบบสอบฉบับที่ 1 ที่ไม่ได้ปรับเทียบ	152
9	โค้งคุณลักษณะข้อสอบ (ICC) ตามโมเดลการตอบสนองข้อสอบ 3 พารามิเตอร์ของข้อสอบ 5 ข้อ จากแบบสอบฉบับที่ 2 ที่ปรับเทียบเข้าสู่สเกลของแบบสอบฉบับที่ 1 ด้วยวิธี Mean and Sigma	152
10	โค้งคุณลักษณะข้อสอบ (ICC) ตามโมเดลการตอบสนองข้อสอบ 3 พารามิเตอร์ของข้อสอบ 5 ข้อ จากแบบสอบฉบับที่ 1 ที่ไม่ได้ปรับเทียบ	153
11	โค้งคุณลักษณะข้อสอบ (ICC) ตามโมเดลการตอบสนองข้อสอบ 3 พารามิเตอร์ของข้อสอบ 5 ข้อ จากแบบสอบฉบับที่ 2 ที่ปรับเทียบเข้าสู่สเกลของแบบสอบฉบับที่ 1 ด้วยวิธี Characteristic Curve	153
12	โค้งคุณลักษณะข้อสอบ (ICC) ตามโมเดลการตอบสนองข้อสอบ 3 พารามิเตอร์ของข้อสอบ 5 ข้อ จากแบบสอบฉบับที่ 1 ที่ไม่ได้ปรับเทียบ	154
13	โค้งคุณลักษณะข้อสอบ (ICC) ตามโมเดลการตอบสนองข้อสอบ 3 พารามิเตอร์ของข้อสอบ 5 ข้อ จากแบบสอบฉบับที่ 2 ที่ปรับเทียบเข้าสู่สเกลของแบบสอบฉบับที่ 1 ด้วยวิธี วิธี Same Scaling Convention	154