

## บทที่ 4

### ผลการวิเคราะห์ข้อมูล

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อพัฒนาเกณฑ์ตัดสินคุณภาพการเปรียบเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ การนำเสนอผลการวิเคราะห์ข้อมูลแยกเสนอเป็น 4 ตอน ตอนแรกเป็นผลการตรวจสอบความถูกต้องและความเป็นเอกมิติของข้อมูลที่จำลองจากโปรแกรม IRTDATA ตอนที่สองเป็นผลการวิเคราะห์จุดตัดแบ่งกลุ่มค่าดัชนีความแตกต่างออกเป็นดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับสูงและดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับต่ำ ตอนที่สามเป็นเกณฑ์ตัดสินคุณภาพการเปรียบเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบที่พัฒนาขึ้น และตอนสุดท้าย เป็นผลการตรวจสอบคุณภาพของเกณฑ์ที่พัฒนาขึ้น

#### ตอนที่ 1 ผลการตรวจสอบความถูกต้องและความเป็นเอกมิติของข้อมูลที่จำลองจากโปรแกรม IRTDATA

การนำเสนอผลการวิเคราะห์ข้อมูลในตอนที่ 1 นี้ แบ่งออกเป็น 2 ส่วน ส่วนแรกเป็นผลการตรวจสอบความถูกต้องของข้อมูลที่ได้จากการจำลองโดยใช้โปรแกรม IRTDATA และส่วนที่ 2 เป็นผลการตรวจสอบความเป็นเอกมิติของข้อมูลที่จำลองจากโปรแกรม IRTDATA

##### 1.1 ผลการตรวจสอบความถูกต้องของข้อมูลที่ได้จากการจำลองโดยใช้โปรแกรม IRTDATA

1.1.1 ตรวจสอบความถูกต้องของข้อมูลจากการจำลองข้อมูลจากโปรแกรม IRTDATA โมเดล 1 พารามิเตอร์ โดยจำลองข้อมูลในเงื่อนไขใช้แบบสอบที่มีจำนวนข้อสอบ 40 ข้อ ผู้สอบจำนวน 300 คน และแบบสอบที่มีจำนวนข้อสอบ 80 ข้อ ผู้สอบจำนวน 1,000 คน นำคะแนนที่ได้จากการจำลองข้อมูลซึ่งเป็นคะแนนดิบของการตอบข้อสอบแต่ละข้อจากการตอบของผู้สอบแต่ละคนไปวิเคราะห์โดยใช้โปรแกรม BILOG พิจารณาจากพารามิเตอร์ผู้ตอบ ( $\theta$ ) และค่าพารามิเตอร์ข้อสอบ ค่าความยาก ( $b$ ) นำไปเปรียบเทียบกับผลที่ได้จากโปรแกรม IRTDATA แล้วได้ทดสอบความแตกต่างระหว่างค่าเฉลี่ยพารามิเตอร์ผู้สอบ ค่าพารามิเตอร์ของข้อสอบที่ได้จากโปรแกรมทั้งสอง โดยใช้ t-test สำหรับทดสอบกลุ่มตัวอย่างสองกลุ่มที่เกี่ยวข้องกัน (Two Dependent Samples Test) ปรากฏผลดังนี้

ตารางที่ 2 ความแตกต่างระหว่างค่าเฉลี่ยพารามิเตอร์ผู้ตอบ 300 คน พารามิเตอร์ข้อสอบ 40 ข้อ จากโปรแกรม IRTDATA และ โปรแกรม BILOG โมเดล 1 พารามิเตอร์

ค่าพารามิเตอร์	$\bar{X}$	S.D.	$\bar{d}$	t	df	$\alpha$
$\theta$ จาก IRTDATA	-0.0140	1.0142	-0.0140	-0.780	299	.436
$\theta$ จาก BILOG	0.0000	0.9747				
b จาก IRTDATA	-0.0433	0.9210	-0.0214	-1.055	39	.298
b จาก BILOG	-0.0219	1.0070				

จากตารางที่ 2 ค่าเฉลี่ยพารามิเตอร์ผู้สอบ 300 คน และค่าเฉลี่ยพารามิเตอร์ข้อสอบ 40 ข้อ จากโปรแกรม IRTDATA และและวิเคราะห์ด้วยโปรแกรม BILOG พบว่าไม่แตกต่างกันอย่างมีนัยสำคัญที่ระดับ .05 สรุปได้ว่าค่าพารามิเตอร์ผู้สอบ  $\theta$  และพารามิเตอร์ข้อสอบ ค่าความยาก b ที่ได้จากโปรแกรม IRTDATA และจากการวิเคราะห์ด้วยโปรแกรม BILOG ไม่แตกต่างกัน แสดงว่าข้อมูลที่จำลองขึ้นด้วยโปรแกรม IRTDATA ตามทฤษฎีการตอบสนอง โมเดล 1 พารามิเตอร์มีความถูกต้อง

ตารางที่ 3 ความแตกต่างระหว่างค่าเฉลี่ยพารามิเตอร์ผู้ตอบ 1,000 คน พารามิเตอร์ข้อสอบ 80 ข้อ จากโปรแกรม IRTDATA และ โปรแกรม BILOG โมเดล 1 พารามิเตอร์

ค่าพารามิเตอร์	$\bar{X}$	S.D.	$\bar{d}$	t	df	$\alpha$
$\theta$ จาก IRTDATA	-0.0125	0.9901	-0.0117	-1.894	999	.058
$\theta$ จาก BILOG	0.0008	0.9859				
b จาก IRTDATA	0.0115	1.2193	-0.0065	-0.728	79	.469
b จาก BILOG	0.0181	1.2734				

จากตารางที่ 3 ค่าเฉลี่ยพารามิเตอร์ผู้สอบ 1,000 คน และค่าเฉลี่ยพารามิเตอร์ข้อสอบ 80 ข้อ จากโปรแกรม IRTDATA และและวิเคราะห์ด้วยโปรแกรม BILOG พบว่าไม่แตกต่างกันอย่างมีนัยสำคัญที่ระดับ .05 สรุปได้ว่าค่าพารามิเตอร์ผู้สอบ  $\theta$  และพารามิเตอร์ข้อสอบ ค่าความยาก b ที่ได้จากโปรแกรม IRTDATA และจากการวิเคราะห์ด้วยโปรแกรม BILOG ไม่แตกต่างกัน แสดงว่าข้อมูลที่จำลองขึ้นด้วยโปรแกรม IRTDATA ตามทฤษฎีการตอบสนอง โมเดล 1 พารามิเตอร์มีความถูกต้อง

1.1.2 ตรวจสอบความถูกต้องของข้อมูลจากการจำลองข้อมูลจากโปรแกรม IRTDATA โมเดล 3 พารามิเตอร์ โดยจำลองข้อมูลในเงื่อนไขใช้แบบสอบที่มีจำนวนข้อสอบ 40 ข้อ ผู้สอบจำนวน 2,000 คน และแบบสอบที่มีจำนวนข้อสอบ 60 ข้อ ผู้สอบ 500 คน นำคะแนนที่ได้จากการจำลองข้อมูลซึ่งเป็นคะแนนดิบของการตอบข้อสอบแต่ละข้อจากการตอบของผู้สอบแต่ละคนไปวิเคราะห์โดยใช้โปรแกรม BILOG พิจารณาจากพารามิเตอร์ผู้ตอบ ( $\theta$ ) ค่าพารามิเตอร์ข้อสอบ ค่าอำนาจจำแนก (a) ค่าความยาก (b) และความน่าจะเป็นในการเดาถูก (c) นำไปเปรียบเทียบกับผลที่ได้จากโปรแกรม IRTDATA แล้วได้ทดสอบความแตกต่างระหว่างค่าเฉลี่ยพารามิเตอร์ผู้สอบ ค่าพารามิเตอร์ของข้อสอบที่ได้จากโปรแกรมทั้งสอง โดยใช้ t-test สำหรับทดสอบกลุ่มตัวอย่างสองกลุ่มที่เกี่ยวข้องกัน (Two Dependent Samples Test) ปรากฏผลดังต่อไปนี้

ตารางที่ 4 ความแตกต่างระหว่างค่าเฉลี่ยพารามิเตอร์ผู้ตอบ 2,000 คน พารามิเตอร์ข้อสอบ 40 ข้อ จากโปรแกรม IRTDATA และ โปรแกรม BILOG โมเดล 3 พารามิเตอร์

ค่าพารามิเตอร์	$\bar{X}$	S.D.	$\bar{d}$	t	df	$\alpha$
$\theta$ จาก IRTDATA	0.0047	1.0091	0.0047	0.690	1999	.490
$\theta$ จาก BILOG	0.0000	0.9609				
a จาก IRTDATA	1.4142	0.4952	0.0831	2.006	39	.052
a จาก BILOG	1.3311	0.4451				
b จาก IRTDATA	-0.1703	1.3382	0.0501	1.702	39	.097
b จาก BILOG	-0.2203	1.3866				
c จาก IRTDATA	0.2595	0.0849	0.0105	1.113	39	.273
c จาก BILOG	0.2490	0.0679				

จากตารางที่ 4 ค่าเฉลี่ยพารามิเตอร์ผู้สอบ 2,000 คน และค่าเฉลี่ยพารามิเตอร์ข้อสอบ 40 ข้อ จากโปรแกรม IRTDATA และวิเคราะห์ด้วยโปรแกรม BILOG พบว่าไม่แตกต่างกันอย่างมีนัยสำคัญที่ระดับ .05 สรุปได้ว่าค่าพารามิเตอร์ผู้สอบ  $\theta$  พารามิเตอร์ข้อสอบ ค่าอำนาจจำแนก (a) ค่าความยาก (b) และความน่าจะเป็นในการเดาถูก (c) ที่ได้จากโปรแกรม IRTDATA และจากการวิเคราะห์ด้วยโปรแกรม BILOG ไม่แตกต่างกัน แสดงว่าข้อมูลที่จำลองขึ้นด้วยโปรแกรม IRTDATA ตามทฤษฎีการตอบสนอง โมเดล 3 พารามิเตอร์มีความถูกต้อง

ตารางที่ 5 ความแตกต่างระหว่างค่าเฉลี่ยพารามิเตอร์ผู้ตอบ 500 คน พารามิเตอร์ข้อสอบ 60 ข้อ จากโปรแกรม IRTDATA และ โปรแกรม BILOG โมเดล 3 พารามิเตอร์

ค่าพารามิเตอร์	$\bar{X}$	S.D.	$\bar{d}$	t	df	$\alpha$
$\theta$ จาก IRTDATA	-0.0186	0.9476	-0.0186	1.722	499	.086
$\theta$ จาก BILOG	0.0000	0.9809				
a จาก IRTDATA	1.5420	0.5747	-0.0255	1.909	59	.061
a จาก BILOG	1.5675	0.5571				
b จาก IRTDATA	-0.2330	0.9376	-0.0059	0.243	59	.809
b จาก BILOG	-0.2270	1.0346				
c จาก IRTDATA	0.2077	0.0833	-0.0030	0.369	59	.714
c จาก BILOG	0.2107	0.0645				

จากตารางที่ 5 ค่าเฉลี่ยพารามิเตอร์ผู้สอบ 500 คน และค่าเฉลี่ยพารามิเตอร์ข้อสอบ 60 ข้อ จากโปรแกรม IRTDATA และวิเคราะห์ด้วยโปรแกรม BILOG พบว่าไม่แตกต่างกันอย่างมีนัยสำคัญที่ระดับ .05 สรุปได้ว่าค่าพารามิเตอร์ผู้สอบ  $\theta$  พารามิเตอร์ข้อสอบ ค่าอำนาจจำแนก (a) ค่าความยาก (b) และความน่าจะเป็นในการเดาถูก (c) ที่ได้จากโปรแกรม IRTDATA และจากการวิเคราะห์ด้วยโปรแกรม BILOG ไม่แตกต่างกัน แสดงว่าข้อมูลที่จำลองขึ้นด้วยโปรแกรม IRTDATA ตามทฤษฎีการตอบสนอง โมเดล 3 พารามิเตอร์มีความถูกต้อง

## 1.2 ผลการตรวจสอบความเป็นเอกมิติของข้อมูลที่จำลองจากโปรแกรม IRTDATA

1.2.1 ตรวจสอบความเป็นเอกมิติของข้อมูล ตามทฤษฎีการตอบสนองข้อสอบ โมเดล 1 พารามิเตอร์ โดยการนำข้อมูลผลการตอบข้อสอบรายชื่อที่จำลองจากโปรแกรม IRTDATA ข้อสอบสอบ 40 ข้อ ผู้สอบ 300 คน และใช้ข้อสอบ 80 ข้อ ผู้สอบ 1,000 คน ไปวิเคราะห์ตัวประกอบ (Factor Analysis) ด้วยวิธีวิเคราะห์ตัวประกอบสำคัญ (Principal Component Analysis) และหมุนแกนด้วยวิธี Varimax ตามที่ Lord (1980) เสนอแนะ ใช้โปรแกรมสำเร็จรูป SPSS ปรากฏผลดังนี้



ตารางที่ 6 ค่าไอเกนและร้อยละของความแปรปรวนของตัวประกอบ สำหรับข้อมูลที่ได้จาก  
โปรแกรม IRTDATA โมเดล 1 พารามิเตอร์ ข้อสอบ 40 ข้อ ผู้สอบ 300 คน

ตัวประกอบที่	ค่าไอเกน	ร้อยละของความแปรปรวน
1	11.712	29.280
2	1.208	3.019
3	0.654	1.635
4	0.509	1.272
5	0.544	1.361
6	0.536	1.341
7	0.497	1.243
8	0.410	1.024

จากตารางที่ 6 พบว่า ค่าไอเกนสูงสุดของแบบสอบคือ 11.712 และคิดเป็นร้อยละ 29.28 ของความแปรปรวนทั้งหมด พิจารณาแล้วเห็นว่ามีความพอที่จะสรุปว่ามีความเป็นเอกมิติประเภทที่มีตัวประกอบหลักตัวแรกเด่นกว่าตัวประกอบอื่นๆ (Essential Unidimensionality) ตามเกณฑ์ของ Rackase (Rackase อ้างใน Raju, 1993) ที่เสนอแนะว่าการใช้ทฤษฎีการตอบสนองข้อสอบวิเคราะห์ข้อสอบนั้น ค่าความแปรปรวนจากตัวประกอบหลักตัวแรกของแบบสอบควรมีค่าไม่ต่ำกว่าร้อยละ 20.00 และอัตราส่วนระหว่างตัวประกอบที่ 1 และตัวประกอบที่ 2 เท่ากับ 9.695 ตามเกณฑ์ของ วรณูช แหยมแสง (2536) ที่เสนอว่าอัตราส่วนระหว่างตัวประกอบที่ 1 และตัวประกอบที่ 2 ไม่ควรต่ำกว่า 5.739 ด้วยเหตุนี้จึงสรุปได้ว่าข้อมูลที่ได้จากการใช้โปรแกรม IRTDATA มีความเป็นเอกมิติเป็นไปตามข้อตกลงของทฤษฎีการตอบสนองข้อสอบ

ตารางที่ 7 ค่าไอเกนและร้อยละของความแปรปรวนของตัวประกอบ สำหรับข้อมูลที่ได้จาก  
โปรแกรม IRTDATA โมเดล 1 พารามิเตอร์ ข้อสอบ 80 ข้อ ผู้สอบ 1,000 คน

ตัวประกอบที่	ค่าไอเกน	ร้อยละของความแปรปรวน
1	21.511	26.889
2	2.969	3.711
3	1.219	1.524
4	1.193	1.492
5	1.144	1.430

จากตารางที่ 7 พบว่า ค่าไอเกนสูงสุดของแบบสอบคือ 21.511 และคิดเป็นร้อยละ 26.889 ของความแปรปรวนทั้งหมด พิจารณาแล้วเห็นว่ามีค่ามากพอที่จะสรุปว่ามีความเป็นเอกมิติประเภทที่มีตัวประกอบหลักตัวแรกเด่นกว่าตัวประกอบอื่น ๆ (Essential Unidimensionality) ตามเกณฑ์ของ Rackase (Rackase อ้างใน Raju, 1993) ที่เสนอแนะว่าการใช้ทฤษฎีการตอบสนองข้อสอบวิเคราะห์ข้อสอบนั้น ค่าความแปรปรวนจากตัวประกอบหลักตัวแรกของแบบสอบควรมีค่าไม่ต่ำกว่าร้อยละ 20.00 และอัตราส่วนระหว่างตัวประกอบที่ 1 และตัวประกอบที่ 2 เท่ากับ 7.245 ตามเกณฑ์ของ วรณัฐ แหยมแสง (2536) ที่เสนอว่าอัตราส่วนระหว่างตัวประกอบที่ 1 และตัวประกอบที่ 2 ไม่ควรต่ำกว่า 5.739 ด้วยเหตุนี้จึงสรุปได้ว่าข้อมูลที่ได้จากการใช้โปรแกรม IRTDATA มีความเป็นเอกมิติเป็นไปตามข้อตกลงของทฤษฎีการตอบสนองข้อสอบ

1.2.2 ตรวจสอบความเป็นเอกมิติของข้อมูล ตามทฤษฎีการตอบสนองข้อสอบ โมเดล 3 พารามิเตอร์ โดยการนำข้อมูลผลการตอบข้อสอบรายข้อที่จำลองจากโปรแกรม IRTDATA ข้อสอบ 40 ข้อ ผู้สอบ 2,000 คน และใช้ข้อสอบ 60 ข้อ ผู้สอบ 500 คน ไปวิเคราะห์ตัวประกอบ (Factor Analysis) ด้วยวิธีวิเคราะห์ตัวประกอบสำคัญ (Principal Component Analysis) และหมุนแกนด้วยวิธี Varimax ตามที่ Lord (1980) เสนอแนะ ใช้โปรแกรมสำเร็จรูป SPSS ปรากฏผลดังนี้

ตารางที่ 8 ค่าไอเกนและร้อยละของความแปรปรวนของตัวประกอบ สำหรับข้อมูลที่ได้จากโปรแกรม IRTDATA โมเดล 3 พารามิเตอร์ ใช้ข้อสอบ 40 ข้อ ผู้สอบ 2,000 คน

ตัวประกอบที่	ค่าไอเกน	ร้อยละของความแปรปรวน
1	8.734	21.834
2	1.131	2.828
3	0.478	1.195
4	0.195	0.487

จากตารางที่ 8 พบว่า ค่าไอเกนสูงสุดของแบบสอบคือ 8.734 และคิดเป็นร้อยละ 21.834 ของความแปรปรวนทั้งหมด พิจารณาแล้วเห็นว่ามีค่ามากพอที่จะสรุปว่ามีความเป็นเอกมิติประเภทที่มีตัวประกอบหลักตัวแรกเด่นกว่าตัวประกอบอื่น ๆ (Essential Unidimensionality) ตามเกณฑ์ของ Rackase (Rackase อ้างใน Raju, 1993) ที่เสนอแนะว่าการใช้ทฤษฎีการตอบสนองข้อสอบวิเคราะห์ข้อสอบนั้น ค่าความแปรปรวนจากตัวประกอบหลักตัวแรกของแบบสอบควรมีค่าไม่ต่ำกว่าร้อยละ 20.00 และอัตราส่วนระหว่างตัวประกอบที่ 1 และ

ตัวประกอบที่ 2 เท่ากับ 7.722 ตามเกณฑ์ของ วรรณุช แหยมแสง (2536) ที่เสนอว่าอัตราส่วนระหว่างตัวประกอบที่ 1 และตัวประกอบที่ 2 ไม่ควรต่ำกว่า 5.739 และด้วยเหตุนี้จึงสรุปได้ว่า ข้อมูลที่ได้จากการใช้โปรแกรม IRTDATA มีความเป็นเอกมิติเป็นไปตามข้อตกลงของทฤษฎีการตอบสนองข้อสอบ

ตารางที่ 9 ค่าไอเกนและร้อยละของความแปรปรวนของตัวประกอบ สำหรับข้อมูลที่ได้จากโปรแกรม IRTDATA โมเดล 3 พารามิเตอร์ ใช้ข้อสอบ 60 ข้อ ผู้สอบ 500 คน

ตัวประกอบที่	ค่าไอเกน	ร้อยละของความแปรปรวน
1	12.616	21.026
2	2.010	4.017
3	1.460	2.433
4	1.395	2.324
5	1.326	2.211

จากตารางที่ 9 พบว่า ค่าไอเกนสูงสุดของแบบสอบคือ 12.616 และคิดเป็นร้อยละ 21.026 ของความแปรปรวนทั้งหมด พิจารณาแล้วเห็นว่ามีความพอที่จะสรุปว่ามีความเป็นเอกมิติประเภทที่มีตัวประกอบหลักตัวแรกเด่นกว่าตัวประกอบอื่นๆ (Essential Unidimensionality) ตามเกณฑ์ของ Rackase (Rackase อ้างใน Raju,1993) ที่เสนอแนะว่าการใช้ทฤษฎีการตอบสนองข้อสอบวิเคราะห์ข้อสอบนั้น ค่าความแปรปรวนจากตัวประกอบหลักตัวแรกของแบบสอบควรมีค่าไม่ต่ำกว่าร้อยละ 20.00 และอัตราส่วนระหว่างตัวประกอบที่ 1 และตัวประกอบที่ 2 เท่ากับ 6.276 ตามเกณฑ์ของ วรรณุช แหยมแสง (2536) ที่เสนอว่าอัตราส่วนระหว่างตัวประกอบที่ 1 และตัวประกอบที่ 2 ไม่ควรต่ำกว่า 5.739 และด้วยเหตุนี้จึงสรุปได้ว่า ข้อมูลที่ได้จากการใช้โปรแกรม IRTDATA มีความเป็นเอกมิติเป็นไปตามข้อตกลงของทฤษฎีการตอบสนองข้อสอบ

## ตอนที่ 2 ผลการวิเคราะห์จุดตัดแบ่งกลุ่มค่าดัชนีความแตกต่างออกเป็นดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับสูงและดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับต่ำ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อพัฒนาเกณฑ์ตัดสินคุณภาพการเปรียบเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ เป็นเกณฑ์ที่กำหนดจากค่าดัชนีความแตกต่างระหว่างคะแนนจริงที่เปรียบเทียบแล้วกับคะแนนจริงที่ไม่ได้เปรียบเทียบสำหรับผู้สอบกลุ่มเดียว และได้วิเคราะห์หาจุดตัด ซึ่งเป็นจุดที่แบ่งดัชนีออกเป็นกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับสูงและกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับต่ำ จุดตัดนี้เป็นค่าดัชนีความแตกต่างที่ได้ทดสอบความแตกต่างระหว่างค่าเฉลี่ยคะแนนจริงที่เปรียบเทียบแล้วกับค่าเฉลี่ยคะแนนจริงที่ไม่ได้เปรียบเทียบด้วยสถิติ t-test แบบ Two Dependent Sample Test เมื่อค่า t จากการคำนวณมีค่าเท่ากับค่าวิกฤต ผลการวิเคราะห์หาจุดตัดดังรายละเอียดต่อไปนี้

ตารางที่ 10 การทดสอบความแตกต่างของค่าเฉลี่ยคะแนนจริงที่ไม่ได้ปรับและคะแนนจริงที่ปรับแล้ว สำหรับกลุ่มผู้สอบกลุ่มเดียวกัน จำนวน 1,200 คน แบบสอบจำนวน 60 ข้อ โมเดล 3 พารามิเตอร์ เปรียบเทียบคะแนนด้วยวิธี Mean and Sigma รูปแบบการเก็บรวบรวมข้อมูลด้วยการใช้ข้อสอบร่วม และหาคุณภาพการเปรียบเทียบโดยใช้กลุ่มสอบทานผล

คะแนนจริง	ค่าเฉลี่ย	ส่วนเบี่ยงเบนมาตรฐาน	ค่าความต่างเฉลี่ย	t	df	$\alpha$
ที่ไม่ได้ปรับ	37.7130	11.5590	0.002690	2.587	1,199	.01
ที่ปรับแล้ว	37.7157	11.5585				

การวิเคราะห์ข้อมูลเป็นการใช้ t-test แบบ Two Dependent Sample Test ทดสอบความแตกต่างค่าเฉลี่ยคะแนนจริงที่ปรับแล้วและค่าเฉลี่ยคะแนนจริงที่ไม่ได้ปรับ ในเงื่อนไขการเปรียบเทียบคะแนนและการหาเกณฑ์ในรูปแบบที่ต่างกัน เพื่อหาค่าวิกฤตโดยกำหนดนัยสำคัญทางสถิติที่ระดับ .01 พบว่า การหาคุณภาพการเปรียบเทียบโดยใช้กลุ่มสอบทานผล รูปแบบการเก็บรวบรวมข้อมูลใช้ข้อสอบร่วม เปรียบเทียบคะแนนด้วยวิธี Mean and Sigma จากตารางที่ 10 ให้ค่าความต่างเฉลี่ยเป็น 0.002690 ค่า  $t = 2.587$   $df = 1,199$  และมีนัยสำคัญทางสถิติเป็น .01 ซึ่งเป็นค่าวิกฤต แล้วนำคะแนนจริงที่ปรับแล้วและคะแนนจริงที่ไม่ได้ปรับไปคำนวณค่าดัชนีความแตกต่าง AMD, MAD และ RMS ที่เป็นจุดตัดแบ่งกลุ่มคุณภาพของดัชนีเป็นกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับสูงและกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับต่ำ ดังตารางที่ 11



ตารางที่ 11 ดัชนีความแตกต่าง AMD, MAD และ RMS ที่กำหนดเป็นจุดตัดแบ่งกลุ่มดัชนี ออกเป็นกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงและกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำ ผลจากตารางที่ 10

ดัชนีความแตกต่าง	ค่าดัชนี
AMD	0.000233
MAD	0.002562
RMS	0.003122

จากตารางที่ 11 เป็นค่าดัชนีความแตกต่างที่กำหนดเป็นจุดตัดแบ่งกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงและกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำ ได้จากการหาคุณภาพการปรับเทียบโดยใช้กลุ่มสอบทานผล ปรับเทียบคะแนนด้วยวิธี Mean and Sigma รูปแบบการเก็บรวบรวมข้อมูลใช้ข้อสอบร่วม โมเดล 3 พารามิเตอร์ จำนวนผู้สอบ 1,200 คน แบบสอบจำนวน 60 ข้อ ได้ดัชนี AMD เท่ากับ 0.000233 ดัชนี MAD เท่ากับ 0.002562 และดัชนี RMS เท่ากับ 0.003122

ตารางที่ 12 การทดสอบความแตกต่างของค่าเฉลี่ยคะแนนจริงที่ไม่ได้ปรับและคะแนนจริงที่ปรับแล้วสำหรับกลุ่มผู้สอบกลุ่มเดียวกัน จำนวน 900 คน แบบสอบจำนวน 40 ข้อ โมเดล 1 พารามิเตอร์ ปรับเทียบคะแนนด้วยวิธี Mean and Sigma รูปแบบการเก็บรวบรวมข้อมูลด้วยการใช้ข้อสอบร่วม และหาคุณภาพการปรับเทียบโดยการปรับเทียบกลับสู่แบบสอบเดิม

คะแนนจริง	ค่าเฉลี่ย	ส่วนเบี่ยงเบน มาตรฐาน	ค่าความต่าง เฉลี่ย	t	df	$\alpha$
ที่ไม่ได้ปรับ	17.0535	9.9077	0.002583	2.577	899	.01
ที่ปรับแล้ว	17.0561	9.9066				

การวิเคราะห์ข้อมูลเป็นการใช้ t-test แบบ Two Dependent Sample Test ทดสอบความแตกต่างค่าเฉลี่ยคะแนนจริงที่ปรับแล้วและค่าเฉลี่ยคะแนนจริงที่ไม่ได้ปรับ ในเงื่อนไขการปรับเทียบคะแนนและการหาเกณฑ์ในรูปแบบที่ต่างกัน เพื่อหาค่าวิกฤตโดยกำหนดนัยสำคัญทางสถิติที่ระดับ .01 พบว่า การหาคุณภาพการปรับเทียบโดยการปรับเทียบกลับสู่แบบสอบเดิม แบบแผนการเก็บรวบรวมข้อมูลใช้ข้อสอบร่วม ปรับเทียบคะแนนด้วยวิธี Mean and Sigma จากตารางที่ 12 ให้ค่าความต่างเฉลี่ยเป็น 0.002583 ค่า  $t = 2.577$   $df = 899$

และมีนัยสำคัญทางสถิติเป็น .01 ซึ่งเป็นค่าวิกฤต แล้วนำคะแนนจริงที่ปรับแล้วและคะแนนจริงที่ไม่ได้ปรับไปคำนวณค่าดัชนีความแตกต่าง AMD, MAD และ RMS ที่เป็นจุดตัดแบ่งกลุ่มคุณภาพของดัชนีเป็นกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูง และกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำ ดังตารางที่ 13

ตารางที่ 13 ดัชนีความแตกต่าง AMD, MAD และ RMS ที่กำหนดเป็นจุดตัดแบ่งกลุ่มดัชนี ออกเป็นกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงและกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำ ผลจากตารางที่ 12

ดัชนีความแตกต่าง	ค่าดัชนี
AMD	0.000261
MAD	0.002508
RMS	0.003045

จากตารางที่ 13 เป็นค่าดัชนีความแตกต่างที่กำหนดเป็นจุดตัดแบ่งกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงและกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำ ได้จากการหาคุณภาพการปรับเทียบโดยการปรับเทียบกลับสู่แบบสอบเดิม ปรับเทียบคะแนนด้วยวิธี Mean and Sigma แบบแผนการเก็บรวบรวมข้อมูลใช้ข้อสอบรวม โมเดล 1 พารามิเตอร์ จำนวนผู้สอบ 900 คน แบบสอบจำนวน 50 ข้อ ได้ดัชนี AMD เท่ากับ 0.000261 ดัชนี MAD เท่ากับ 0.002508 และดัชนี RMS เท่ากับ 0.003045

ตารางที่ 14 การทดสอบความแตกต่างของค่าเฉลี่ยคะแนนจริงที่ไม่ได้ปรับและคะแนนจริงที่ปรับแล้ว สำหรับกลุ่มผู้สอบกลุ่มเดียวกัน จำนวน 1,000 คน แบบสอบจำนวน 30 ข้อ โมเดล 3 พารามิเตอร์ ปรับเทียบคะแนนด้วยวิธีกำหนดค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าความสามารถผู้สอบให้เท่ากันก่อนการวิเคราะห์ แบบแผนการเก็บรวบรวมข้อมูลด้วยการใช้กลุ่มสมมูล และหาคุณภาพการปรับเทียบโดยใช้กลุ่มสอบทานผล

คะแนนจริง	ค่าเฉลี่ย	ส่วนเบี่ยงเบน มาตรฐาน	ค่าความต่าง เฉลี่ย	t	df	$\alpha$
ที่ไม่ได้ปรับ	20.2447	6.1824	0.001465	2.565	999	.01
ที่ปรับแล้ว	20.2461	6.1800				

การวิเคราะห์ข้อมูลเป็นการใช้ t-test แบบ Two Dependent Sample Test ทดสอบความแตกต่างค่าเฉลี่ยคะแนนจริงที่ปรับแล้วและค่าเฉลี่ยคะแนนจริงที่ไม่ได้ปรับ ในเงื่อนไขการปรับเทียบคะแนนและการหาเกณฑ์ในรูปแบบที่ต่างกัน เพื่อหาค่าวิกฤตโดยกำหนดนัยสำคัญทางสถิติที่ระดับ .01 พบว่า การหาคุณภาพการปรับเทียบโดยใช้กลุ่มสอบทานผล แบบแผนการเก็บรวบรวมข้อมูลใช้กลุ่มสมมูล ปรับเทียบคะแนนด้วยวิธีกำหนดค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าความสามารถผู้สอบให้เท่ากันก่อนการวิเคราะห์ จากตารางที่ 14 ให้ค่าความต่างเฉลี่ยเป็น 0.001465 ค่า  $t = 2.565$   $df = 999$  และมีนัยสำคัญทางสถิติเป็น .01 ซึ่งเป็นค่าวิกฤต แล้วนำคะแนนจริงที่ปรับแล้วและคะแนนจริงที่ไม่ได้ปรับไปคำนวณค่าดัชนีความแตกต่าง AMD, MAD และ RMS ที่เป็นจุดตัดแบ่งกลุ่มคุณภาพของดัชนีเป็นกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงและกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำ ดังตารางที่ 15

ตารางที่ 15 ดัชนีความแตกต่าง AMD, MAD และ RMS ที่กำหนดเป็นจุดตัดแบ่งกลุ่มดัชนี ออกเป็นกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงและกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำ ผลจากตารางที่ 14

ดัชนีความแตกต่าง	ค่าดัชนี
AMD	0.000237
MAD	0.002375
RMS	0.002930

จากตารางที่ 15 เป็นค่าดัชนีความแตกต่างที่กำหนดเป็นจุดตัดแบ่งกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงและกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำ ได้จากการหาคุณภาพการปรับเทียบโดยใช้กลุ่มสอบทานผล ปรับเทียบคะแนนด้วยวิธีกำหนดค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าความสามารถผู้สอบให้เท่ากันก่อนการวิเคราะห์ แบบแผนการเก็บรวบรวมข้อมูลใช้กลุ่มสมมูล โมเดล 3 พารามิเตอร์ จำนวนผู้สอบ 1,000 คน แบบสอบจำนวน 30 ข้อ ได้ดัชนี AMD เท่ากับ 0.000234 ดัชนี MAD เท่ากับ 0.002375 และดัชนี RMS เท่ากับ 0.002930

จากการปรับเทียบคะแนนที่หลากหลายเงื่อนไขทำให้ได้ค่าที่เป็นจุดตัดแบ่งดัชนีที่แสดงคุณภาพการปรับเทียบคะแนนหลายค่า ดังนั้นจึงได้เฉลี่ยค่าดัชนีและจัดเป็นจุดตัดตามเงื่อนไขการปรับเทียบคะแนนดังต่อไปนี้

ตารางที่ 16 ดัชนีความแตกต่าง AMD, MAD และ RMS ที่กำหนดเป็นจุดตัดแบ่งกลุ่มดัชนี ออกเป็นกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงและกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำ จำแนกตามเงื่อนไขการปรับเทียบคะแนน

เงื่อนไขการปรับเทียบคะแนน	ค่าดัชนีความแตกต่าง		
	AMD	MAD	RMS
รวมเงื่อนไขทั้งหมด	0.000244	0.002481	0.003032
โมเดล 1 พารามิเตอร์	0.000261	0.002508	0.003045
โมเดล 3 พารามิเตอร์	0.000235	0.002468	0.003026
แบบแผนกลุ่มสมมูล	0.000237	0.002375	0.002930
แบบแผนใช้ข้อสอบรวม	0.000247	0.002535	0.003083
การปรับเทียบกลับสู่แบบสอบเดิม	0.000261	0.002508	0.003045
การใช้กลุ่มสอบทานผล	0.000235	0.002468	0.003026

จากตารางที่ 16 พบว่าค่าดัชนีความแตกต่างที่กำหนดเป็นจุดตัดแบ่งกลุ่มดัชนีออกเป็นกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงกับกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำแตกต่างกันตามเงื่อนไขการปรับเทียบคะแนน โดยดัชนี AMD เงื่อนไขการปรับเทียบกลับสู่แบบสอบเดิมและโมเดล 1 พารามิเตอร์มีค่าสูงสุด เท่ากับ 0.000261 รองลงมาคือเงื่อนไขแบบแผนใช้ข้อสอบรวม รวมเงื่อนไขทั้งหมด แบบแผนกลุ่มสมมูล ให้ค่าดัชนีเท่ากับ 0.000247, 0.000244 และ 0.000237 ตามลำดับ ส่วนเงื่อนไขโมเดล 3 พารามิเตอร์และการใช้กลุ่มสอบทานผลมีค่าดัชนีเท่ากันคือ 0.000235

ดัชนี MAD เงื่อนไขแบบแผนใช้ข้อสอบรวมมีค่าสูงสุด เท่ากับ 0.002535 รองลงมาคือเงื่อนไขโมเดล 1 พารามิเตอร์และการปรับเทียบกลับสู่แบบสอบเดิม ให้ค่าดัชนีเท่ากับ 0.002508 รวมเงื่อนไขทั้งหมดให้ค่าดัชนีเท่ากับ 0.002481 โมเดล 3 พารามิเตอร์และการใช้กลุ่มสอบทานผลให้ค่าดัชนีเท่ากันคือ 0.002468 และเงื่อนไขแบบแผนกลุ่มสมมูลให้ค่าดัชนีเท่ากับกับ 0.002375

ดัชนี RMS เงื่อนไขใช้แบบสอบรวมให้ค่าดัชนีสูงสุดเท่ากับ 0.003083 รองลงมาโมเดล 1 พารามิเตอร์และการปรับเทียบกลับสู่แบบสอบเดิมให้ค่าดัชนีเท่ากับ 0.003045 รวมเงื่อนไขทั้งหมดให้ค่าดัชนีเท่ากับ 0.003032 โมเดล 3 พารามิเตอร์และการใช้กลุ่มสอบทานผลให้ค่าดัชนีเท่ากันคือ 0.003026 และการใช้กลุ่มสมมูลให้ค่าดัชนีเท่ากับ 0.002930

### ตอนที่ 3 เกณฑ์ตัดสินคุณภาพการเปรียบเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ

การวิจัยครั้งนี้เป็นการพัฒนาเกณฑ์จากเงื่อนไขที่เป็นไปได้เกือบทั้งหมดของการเปรียบเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ เช่น โมเดลการตอบสนองข้อสอบ การเก็บรวบรวมข้อมูล แบบแผนการหาคุณภาพการเปรียบเทียบ วิธีการเปรียบเทียบคะแนน จำนวนผู้สอบ และจำนวนข้อสอบ ซึ่งทำให้เกิดผลการเปรียบเทียบที่หลากหลาย แต่เงื่อนไขที่สำคัญและอิทธิพลอย่างมากต่อการเปรียบเทียบคะแนนและเป็นเงื่อนไขที่นิยมใช้กันมากคือ โมเดลการตอบสนองข้อสอบ การเก็บรวบรวมข้อมูล และแบบแผนการหาคุณภาพการเปรียบเทียบ เพื่อให้ได้เกณฑ์ที่หลากหลายเป็นทางเลือกสำหรับผู้ใช้งานผลการเปรียบเทียบคะแนน ผู้วิจัยจึงจัดดัชนีทั้งในกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับสูงและกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับต่ำให้เป็นเกณฑ์ตัดสินคุณภาพการเปรียบเทียบซึ่งมี 2 ระดับเป็นเกณฑ์ที่ให้สารสนเทศในการตัดสินผลการเปรียบเทียบคะแนนได้น้อย ประกอบกับการพิจารณาจุดตัดที่เป็นค่าดัชนีความแตกต่างที่ได้ทดสอบความแตกต่างระหว่างค่าเฉลี่ยคะแนนจริงที่เปรียบเทียบแล้วกับค่าเฉลี่ยคะแนนจริงที่ยังไม่ได้เปรียบเทียบ ด้วยสถิติ t-test แบบ Two Dependent Sample Test เมื่อค่า t จากการคำนวณมีค่าเท่ากับค่าวิกฤตที่นัยสำคัญทางสถิติที่ระดับ .01 มีโอกาสที่จะเกิดความคลาดเคลื่อนได้ และโอกาสที่ค่า t คำนวณได้มีค่าเท่ากับค่าวิกฤตที่นัยสำคัญทางสถิติที่ระดับ .02 จะเกิดขึ้นเมื่อผลการทดสอบความแตกต่างระหว่างค่าเฉลี่ยของคะแนนทั้งสองไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ในทำนองเดียวกันค่าที่ที่คำนวณได้มีค่าเท่ากับค่าวิกฤตที่นัยสำคัญทางสถิติที่ระดับ .001 จะเกิดขึ้นเมื่อผลการทดสอบความแตกต่างระหว่างค่าเฉลี่ยของคะแนนทั้งสองแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ผู้วิจัยจึงกำหนดจุดตัดซึ่งเป็นค่าดัชนีความแตกต่างที่ได้ทดสอบความแตกต่างระหว่างคะแนนทั้งสองด้วยสถิติ t-test เมื่อค่า t คำนวณมีค่าเท่ากับค่าวิกฤตที่นัยสำคัญทางสถิติระดับ .02 สำหรับกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับสูง และกำหนดจุดตัดซึ่งเป็นค่าดัชนีความแตกต่างที่ได้ทดสอบความแตกต่างระหว่างคะแนนทั้งสองด้วยสถิติ t-test เมื่อค่า t คำนวณมีค่าเท่ากับค่าวิกฤตที่นัยสำคัญทางสถิติระดับ .001 สำหรับกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับต่ำ เกณฑ์ตัดสินคุณภาพการเปรียบเทียบคะแนนที่ได้แต่ละเกณฑ์จะมี 4 ระดับ ผลการวิเคราะห์ในตอนนี้นำออกเป็น 7 ส่วน ส่วนแรกเป็นจุดตัดที่แบ่งกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับสูงและกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับต่ำออกเป็นสองกลุ่มย่อย ส่วนที่ 2 เป็นเกณฑ์ตัดสินคุณภาพการเปรียบเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ เกณฑ์รวมทุกเงื่อนไข ส่วนที่ 3 เป็นเกณฑ์สำหรับโมเดล 1 พารามิเตอร์ ส่วนที่ 4 เป็นเกณฑ์สำหรับโมเดล 3 พารามิเตอร์ ส่วนที่ 5 เป็นเกณฑ์สำหรับแบบแผนการเก็บรวบรวมข้อมูลเมื่อใช้กลุ่มสมมูล ส่วนที่ 6 เป็นเกณฑ์สำหรับแบบแผนการเก็บรวบรวมข้อมูลเมื่อใช้แบบสอบร่วม ส่วนที่ 7 เป็นเกณฑ์สำหรับการหาคุณภาพการเปรียบเทียบคะแนนเมื่อใช้การปรับเทียบกลับสู่แบบสอบเดิม และส่วนที่ 8 เป็นเกณฑ์สำหรับการหาคุณภาพการเปรียบเทียบเมื่อใช้กลุ่มสอบทานผล

### 3.1 จุดตัดที่แบ่งกลุ่มดัชนีที่แสดงคุณภาพเปรียบเทียบระดับสูงและกลุ่มดัชนีคุณภาพการเปรียบเทียบระดับต่ำออกเป็นสองกลุ่มย่อย

เกณฑ์ที่พัฒนาขึ้นกำหนดจากค่าดัชนีความแตกต่างระหว่างคะแนนจริงที่ปรับเทียบแล้วกับคะแนนจริงที่ไม่ได้ปรับเทียบสำหรับผู้สอบกลุ่มเดียวกัน เมื่อทดสอบความแตกต่างระหว่างค่าเฉลี่ยคะแนนจริงทั้งสองด้วยสถิติ t-test แบบ Dependent Sample Test พบว่าค่า t คำนวณมีค่าเท่ากับค่าวิกฤตที่นัยสำคัญทางสถิติระดับ .02 สำหรับจุดตัดกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับสูง และสำหรับค่า t คำนวณมีค่าเท่ากับค่าวิกฤตที่นัยสำคัญทางสถิติระดับ .001 สำหรับจุดตัดกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับต่ำ จุดตัดที่ได้ตั้งรายละเอียดต่อไปนี้

ตารางที่ 17 การทดสอบความแตกต่างของค่าเฉลี่ยคะแนนจริงที่ปรับและคะแนนจริงที่ปรับแล้ว สำหรับกลุ่มผู้สอบกลุ่มเดียวกัน จำนวน 1,400 คน แบบสอบจำนวน 80 ข้อ โมเดล 1 พารามิเตอร์ เปรียบเทียบคะแนนด้วยวิธี Mean and Sigma รูปแบบการเก็บรวบรวมข้อมูลด้วยการใช้ข้อสอบร่วม และหาคุณภาพการเปรียบเทียบโดยใช้กลุ่มสอบทานผล

คะแนนจริง	ค่าเฉลี่ย	ส่วนเบี่ยงเบน มาตรฐาน	ค่าความต่าง เฉลี่ย	t	df	$\alpha$
ที่ไม่ได้ปรับ	39.2546	19.3251	0.003518	2.333	1,399	.02
ที่ปรับแล้ว	39.2511	19.3249				

การวิเคราะห์ข้อมูลเป็นการใช้ t-test แบบ Two Dependent Sample Test ทดสอบความแตกต่างค่าเฉลี่ยคะแนนจริงที่ปรับแล้วและค่าเฉลี่ยคะแนนจริงที่ไม่ได้ปรับ ในเงื่อนไขการปรับเทียบคะแนนและการหาเกณฑ์ในรูปแบบที่ต่างกัน เพื่อหาค่าวิกฤตโดยกำหนดนัยสำคัญทางสถิติที่ระดับ .02 พบว่า การหาคุณภาพการเปรียบเทียบโดยใช้กลุ่มสอบทานผล แบบแผนการเก็บรวบรวมข้อมูลใช้ข้อสอบร่วม เปรียบเทียบคะแนนด้วยวิธี Mean and Sigma จากตารางที่ 13 ให้ค่าความต่างเฉลี่ยเป็น 0.003518 ค่า  $t = 2.333$   $df = 1,399$  และมีนัยสำคัญทางสถิติเป็น .02 ซึ่งเป็นค่าวิกฤต แล้วนำคะแนนจริงที่ปรับแล้วและคะแนนจริงที่ไม่ได้ปรับไปคำนวณค่าดัชนีความแตกต่าง AMD, MAD และ RMS ที่เป็นจุดตัดแบ่งกลุ่มคุณภาพของดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับสูงออกเป็น 2 กลุ่มย่อย ดังตารางที่ 18

ตารางที่ 18 ดัชนีความแตกต่าง AMD, MAD และ RMS ที่กำหนดเป็นจุดตัดแบ่งกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงออกเป็น 2 กลุ่มย่อย ผลจากตารางที่ 17

ดัชนีความแตกต่าง	ค่าดัชนี
AMD	0.000182
MAD	0.002450
RMS	0.002924

จากตารางที่ 18 เป็นค่าดัชนีความแตกต่างที่กำหนดเป็นจุดตัดแบ่งกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงออกเป็น 2 กลุ่มย่อย ได้จากการหาคุณภาพการปรับเทียบโดยใช้กลุ่มสอบทานผล ปรับเทียบคะแนนด้วยวิธี Mean and Sigma รูปแบบการเก็บรวบรวมข้อมูลใช้ข้อสอบรวม โมเดล 1 พารามิเตอร์ จำนวนผู้สอบ 1,400 คน แบบสอบจำนวน 80 ข้อ ได้ดัชนี AMD เท่ากับ 0.000182 ดัชนี MAD เท่ากับ 0.002450 และดัชนี RMS เท่ากับ 0.002924

ตารางที่ 19 การทดสอบความแตกต่างของค่าเฉลี่ยคะแนนจริงที่ปรับและคะแนนจริงที่ปรับแล้ว สำหรับกลุ่มผู้สอบกลุ่มเดียวกัน จำนวน 1,600 คน แบบสอบจำนวน 90 ข้อ โมเดล 3 พารามิเตอร์ ปรับเทียบคะแนนด้วยวิธี Mean and Sigma รูปแบบการเก็บรวบรวมข้อมูลด้วยการใช้ข้อสอบรวม และหาคุณภาพการปรับเทียบโดยการปรับเทียบกลับสู่แบบสอบเดิม

คะแนนจริง	ค่าเฉลี่ย	ส่วนเบี่ยงเบน มาตรฐาน	ค่าความต่าง เฉลี่ย	t	df	$\alpha$
ที่ไม่ได้ปรับ	59.0914	17.2817	0.002923	2.356	1,599	.02
ที่ปรับแล้ว	59.0943	17.2845				

การวิเคราะห์ข้อมูลเป็นการใช้ t-test แบบ Two Dependent Sample Test ทดสอบความแตกต่างค่าเฉลี่ยคะแนนจริงที่ปรับแล้วและค่าเฉลี่ยคะแนนจริงที่ไม่ได้ปรับ ในเงื่อนไขการปรับเทียบคะแนนและการหาเกณฑ์ในรูปแบบที่ต่างกัน เพื่อหาค่าวิกฤตโดยกำหนดนัยสำคัญทางสถิติที่ระดับ .02 พบว่า การหาคุณภาพการปรับเทียบโดยการปรับเทียบกลับสู่แบบสอบเดิม แบบแผนการเก็บรวบรวมข้อมูลใช้ข้อสอบรวม ปรับเทียบคะแนนด้วยวิธี Mean

and Sigma จากตารางที่ 19 ให้ค่าความต่างเฉลี่ยเป็น 0.002923 ค่า  $t = 2.356$   $df = 1,599$  และมีนัยสำคัญทางสถิติเป็น .02 ซึ่งเป็นค่าค่าวิกฤต แล้วนำคะแนนจริงที่ปรับแล้วและคะแนนจริงที่ไม่ได้ปรับไปคำนวณค่าดัชนีความแตกต่าง AMD, MAD และ RMS ที่เป็นจุดตัดแบ่งกลุ่มคุณภาพของดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับสูงออกเป็น 2 กลุ่มย่อย ดังตารางที่ 20

ตารางที่ 20 ดัชนีความแตกต่าง AMD, MAD และ RMS ที่กำหนดเป็นจุดตัดแบ่งกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับสูงออกเป็น 2 กลุ่มย่อย ผลจากตารางที่ 19

ดัชนีความแตกต่าง	ค่าดัชนี
AMD	0.000169
MAD	0.002370
RMS	0.002874

จากตารางที่ 20 เป็นค่าดัชนีความแตกต่างที่กำหนดเป็นจุดตัดแบ่งกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับสูงออกเป็น 2 กลุ่มย่อย ได้จากการหาคุณภาพการเปรียบเทียบโดยการเปรียบเทียบกลับสู่แบบสอบถาม ปรับเทียบคะแนนด้วยวิธี Mean and Sigma แบบแผนการเก็บรวบรวมข้อมูลใช้ข้อสอบรวม โมเดล 3 พารามิเตอร์ จำนวนผู้สอบ 1,600 คน แบบสอบจำนวน 90 ข้อ ได้ดัชนี AMD เท่ากับ 0.000169 ดัชนี MAD เท่ากับ 0.002370 และดัชนี RMS เท่ากับ 0.002874

ตารางที่ 21 การทดสอบความแตกต่างของค่าเฉลี่ยคะแนนจริงที่ปรับและคะแนนจริงที่ปรับแล้ว สำหรับกลุ่มผู้สอบกลุ่มเดียวกัน จำนวน 1,200 คน แบบสอบจำนวน 70 ข้อ โมเดล 3 พารามิเตอร์ ปรับเทียบคะแนนด้วยวิธีกำหนดค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าความสามารถผู้สอบให้เท่ากันก่อนการวิเคราะห์ แบบแผนการเก็บรวบรวมข้อมูลด้วยการใช้กลุ่มสมมูล และหาคุณภาพการเปรียบเทียบโดยใช้กลุ่มสอบทานผล

คะแนนจริง	ค่าเฉลี่ย	ส่วนเบี่ยงเบน มาตรฐาน	ค่าความต่าง เฉลี่ย	t	df	$\alpha$
ที่ไม่ได้ปรับ	44.0739	14.7956	0.002912	2.336	1,199	.02
ที่ปรับแล้ว	44.0710	14.7894				



การวิเคราะห์ข้อมูลเป็นการใช้ t-test แบบ Two Dependent Sample Test ทดสอบความแตกต่างค่าเฉลี่ยคะแนนจริงที่ปรับแล้วและค่าเฉลี่ยคะแนนจริงที่ไม่ได้ปรับ ในเงื่อนไขการปรับเทียบคะแนนและการหาเกณฑ์ในรูปแบบที่ต่างกัน เพื่อหาค่าวิกฤตโดยกำหนดนัยสำคัญทางสถิติที่ระดับ .02 พบว่า การหาคุณภาพการปรับเทียบโดยใช้กลุ่มสอบทานผล แบบแผนการเก็บรวบรวมข้อมูลใช้กลุ่มสมมูล ปรับเทียบคะแนนด้วยวิธีกำหนดค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าความสามารถผู้สอบให้เท่ากันก่อนการวิเคราะห์ จากตารางที่ 21 ให้ค่าความต่างเฉลี่ยเป็น 0.002912 ค่า  $t = 2.336$   $df = 1,199$  และมีนัยสำคัญทางสถิติเป็น .02 ซึ่งเป็นค่าวิกฤต แล้วนำคะแนนจริงที่ปรับแล้วและคะแนนจริงที่ไม่ได้ปรับไปคำนวณค่าดัชนีความแตกต่าง AMD, MAD และ RMS ที่เป็นจุดตัดแบ่งกลุ่มคุณภาพของดัชนีกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงออกเป็น 2 กลุ่มย่อย ดังตารางที่ 22

ตารางที่ 22 ดัชนีความแตกต่าง AMD, MAD และ RMS ที่กำหนดเป็นจุดตัดแบ่งกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงออกเป็น 2 กลุ่มย่อย จากตารางที่ 21

ดัชนีความแตกต่าง	ค่าดัชนี
AMD	0.000197
MAD	0.002353
RMS	0.002924

จากตารางที่ 22 เป็นค่าดัชนีความแตกต่างที่กำหนดเป็นจุดตัดแบ่งกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงออกเป็น 2 กลุ่มย่อย ได้จากการหาคุณภาพการปรับเทียบโดยใช้กลุ่มสอบทานผล ปรับเทียบคะแนนด้วยวิธีกำหนดค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าความสามารถผู้สอบให้เท่ากันก่อนการวิเคราะห์ แบบแผนการเก็บรวบรวมข้อมูลใช้กลุ่มสมมูล โมเดล 3 พารามิเตอร์ จำนวนผู้สอบ 1,200 คนแบบสอบจำนวน 70 ข้อ ได้ดัชนี AMD เท่ากับ 0.000197 ดัชนี MAD เท่ากับ 0.002353 และดัชนี RMS เท่ากับ 0.002924

ตารางที่ 23 การทดสอบความแตกต่างของค่าเฉลี่ยคะแนนจริงที่ปรับและคะแนนจริงที่ปรับแล้ว สำหรับกลุ่มผู้สอบกลุ่มเดียวกัน จำนวน 1,100 คน แบบสอบจำนวน 50 ข้อ โมเดล 3 พารามิเตอร์ ปรับเทียบคะแนนด้วยวิธี Mean and Mean รูปแบบการเก็บรวบรวมข้อมูลด้วยการใช้ข้อสอบร่วม และหาคุณภาพการปรับเทียบโดยใช้กลุ่มสอบทานผล

คะแนนจริง	ค่าเฉลี่ย	ส่วนเบี่ยงเบน มาตรฐาน	ค่าความต่าง เฉลี่ย	t	df	$\alpha$
ที่ไม่ได้ปรับ	34.8338	9.4015	0.003266	3.407	1,099	.001
ที่ปรับแล้ว	34.8306	9.4084				

การวิเคราะห์ข้อมูลเป็นการใช้ t-test แบบ Two Dependent Sample Test ทดสอบความแตกต่างค่าเฉลี่ยคะแนนจริงที่ปรับแล้วและค่าเฉลี่ยคะแนนจริงที่ไม่ได้ปรับ ในเงื่อนไขการปรับเทียบคะแนนและการหาเกณฑ์ในรูปแบบที่ต่างกัน เพื่อหาค่าวิกฤตโดยกำหนดนัยสำคัญทางสถิติที่ระดับ .001 พบว่า การหาคุณภาพการปรับเทียบโดยใช้กลุ่มสอบทานผล แบบแผนการเก็บรวบรวมข้อมูลใช้ข้อสอบร่วม ปรับเทียบคะแนนด้วยวิธี Mean and Mean จากตารางที่ 23 ให้ค่าความต่างเฉลี่ยเป็น 0.003266 ค่า  $t = 3.407$   $df = 1,099$  และมีนัยสำคัญทางสถิติเป็น .001 ซึ่งเป็นค่าวิกฤต แล้วนำคะแนนจริงที่ปรับแล้วและคะแนนจริงที่ไม่ได้ปรับไปคำนวณค่าดัชนีความแตกต่าง AMD, MAD และ RMS ที่เป็นจุดตัดแบ่งกลุ่มคุณภาพของดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำออกเป็น 2 กลุ่มย่อย ดังตารางที่ 24

ตารางที่ 24 ดัชนีความแตกต่าง AMD, MAD และ RMS ที่กำหนดเป็นจุดตัดแบ่งกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำออกเป็น 2 กลุ่มย่อย ผลจากตารางที่ 23

ดัชนีความแตกต่าง	ค่าดัชนี
AMD	0.000347
MAD	0.002717
RMS	0.003398

จากตารางที่ 24 เป็นค่าดัชนีความแตกต่างที่กำหนดเป็นจุดตัดแบ่งกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำออกเป็น 2 กลุ่มย่อย ได้จากการหาคุณภาพการปรับเทียบโดยใช้กลุ่มสอบทานผล ปรับเทียบคะแนนด้วยวิธี Mean and Mean รูปแบบการเก็บรวบรวมข้อมูลใช้

ข้อสอบร่วม โมเดล 3 พารามิเตอร์ จำนวนผู้สอบ 1,100 คน แบบสอบจำนวน 50 ข้อ ได้ดัชนี AMD เท่ากับ 0.000347 ดัชนี MAD เท่ากับ 0.002717 และดัชนี RMS เท่ากับ 0.003398

ตารางที่ 25 การทดสอบความแตกต่างของค่าเฉลี่ยคะแนนจริงที่ปรับและคะแนนจริงที่ปรับแล้ว สำหรับกลุ่มผู้สอบกลุ่มเดียวกัน จำนวน 1,000 คน แบบสอบจำนวน 60 ข้อ โมเดล 1 พารามิเตอร์ ปรับเทียบคะแนนด้วยวิธี Mean and Mean รูปแบบการเก็บรวบรวมข้อมูลด้วยการใช้ข้อสอบร่วม และหาคุณภาพการปรับเทียบโดยการปรับเทียบกลับสู่แบบสอบเดิม

คะแนนจริง	ค่าเฉลี่ย	ส่วนเบี่ยงเบน มาตรฐาน	ค่าความต่าง เฉลี่ย	t	df	$\alpha$
ที่ไม่ได้ปรับ	29.0975	13.5211	0.004538	3.478	999	.001
ที่ปรับแล้ว	29.1020	13.5237				

การวิเคราะห์ข้อมูลเป็นการใช้ t-test แบบ Two Dependent Sample Test ทดสอบความแตกต่างค่าเฉลี่ยคะแนนจริงที่ปรับแล้วและค่าเฉลี่ยคะแนนจริงที่ไม่ได้ปรับ ในเงื่อนไขการปรับเทียบคะแนนและการหาเกณฑ์ในรูปแบบที่ต่างกัน เพื่อหาค่าวิกฤตโดยกำหนดนัยสำคัญทางสถิติที่ระดับ .001 พบว่า การหาคุณภาพการปรับเทียบโดยการปรับเทียบกลับสู่แบบสอบเดิม แบบแผนการเก็บรวบรวมข้อมูลใช้ข้อสอบร่วม ปรับเทียบคะแนนด้วยวิธี Mean and Mean จากตารางที่ 25 ให้ค่าความต่างเฉลี่ยเป็น 0.004538 ค่า  $t = 3.478$   $df = 999$  และมีนัยสำคัญทางสถิติเป็น .001 ซึ่งเป็นค่าค่าวิกฤต แล้วย่นำคะแนนจริงที่ปรับแล้วและคะแนนจริงที่ไม่ได้ปรับไปคำนวณค่าดัชนีความแตกต่าง AMD, MAD และ RMS ที่เป็นจุดตัดแบ่งกลุ่มคุณภาพของดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำออกเป็น 2 กลุ่มย่อย ดังตารางที่ 26

ตารางที่ 26 ดัชนีความแตกต่าง AMD, MAD และ RMS ที่กำหนดเป็นจุดตัดแบ่งกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำออกเป็น 2 กลุ่มย่อย ผลจากตารางที่ 25

ดัชนีความแตกต่าง	ค่าดัชนี
AMD	0.000336
MAD	0.002637
RMS	0.003124

จากตารางที่ 26 เป็นค่าดัชนีความแตกต่างที่กำหนดเป็นจุดตัดแบ่งกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับต่ำออกเป็น 2 กลุ่มย่อย ได้จากการหาคุณภาพการเปรียบเทียบโดยการเปรียบเทียบกลับสู่แบบสอบถามเดิม เปรียบเทียบคะแนนด้วยวิธี Mean and Mean แบบแผนการเก็บรวบรวมข้อมูลใช้ข้อสอบรวม โมเดล 1 พารามิเตอร์ จำนวนผู้สอบ 1,000 คน แบบสอบจำนวน 60 ข้อ ได้ดัชนี AMD เท่ากับ 0.000336 ดัชนี MAD เท่ากับ 0.002637 และดัชนี RMS เท่ากับ 0.003124

ตารางที่ 27 การทดสอบความแตกต่างของค่าเฉลี่ยคะแนนจริงที่ปรับและคะแนนจริงที่ปรับแล้ว สำหรับกลุ่มผู้สอบกลุ่มเดียวกัน จำนวน 700 คน แบบสอบจำนวน 40 ข้อ โมเดล 1 พารามิเตอร์ เปรียบเทียบคะแนนด้วยวิธีกำหนดค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าความสามารถผู้สอบให้เท่ากันก่อนการวิเคราะห์ แบบแผนการเก็บรวบรวมข้อมูลด้วยการใช้กลุ่มสมมูล และหาคุณภาพการเปรียบเทียบโดยใช้กลุ่มสอบทานผล

คะแนนจริง	ค่าเฉลี่ย	ส่วนเบี่ยงเบน มาตรฐาน	ค่าความต่าง เฉลี่ย	t	df	$\alpha$
ที่ไม่ได้ปรับ	20.8937	9.4692	0.003540	3.296	699	.001
ที่ปรับแล้ว	20.8902	9.4677				

การวิเคราะห์ข้อมูลเป็นการใช้ t-test แบบ Two Dependent Sample Test ทดสอบความแตกต่างค่าเฉลี่ยคะแนนจริงที่ปรับแล้วและค่าเฉลี่ยคะแนนจริงที่ไม่ได้ปรับ ในเงื่อนไขการเปรียบเทียบคะแนนและการหาเกณฑ์ในรูปแบบที่ต่างกัน เพื่อหาค่าวิกฤตโดยกำหนดนัยสำคัญทางสถิติที่ระดับ .001 พบว่า การหาคุณภาพการเปรียบเทียบโดยใช้กลุ่มสอบทานผล แบบแผนการเก็บรวบรวมข้อมูลใช้กลุ่มสมมูล เปรียบเทียบคะแนนด้วยวิธีกำหนดค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าความสามารถผู้สอบให้เท่ากันก่อนการวิเคราะห์ จากตารางที่ 27 ให้ค่าความต่างเฉลี่ยเป็น 0.003540 ค่า  $t = 3.296$   $df = 699$  และมีนัยสำคัญทางสถิติเป็น .001 ซึ่งเป็นค่าวิกฤต แล้วนำคะแนนจริงที่ปรับแล้วและคะแนนจริงที่ไม่ได้ปรับไปคำนวณค่าดัชนีความแตกต่าง AMD, MAD และ RMS ที่เป็นจุดตัดแบ่งกลุ่มคุณภาพของดัชนีกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับต่ำออกเป็น 2 กลุ่มย่อย ดังตารางที่ 28

ตารางที่ 28 ดัชนีความแตกต่าง AMD, MAD และ RMS ที่กำหนดเป็นจุดตัดแบ่งกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับต่ำออกเป็น 2 กลุ่มย่อย ผลจากตารางที่ 27

ดัชนีความแตกต่าง	ค่าดัชนี
AMD	0.000374
MAD	0.002510
RMS	0.003022

จากตารางที่ 28 เป็นค่าดัชนีความแตกต่างที่กำหนดเป็นจุดตัดแบ่งกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับต่ำออกเป็น 2 กลุ่มย่อย ได้จากการหาคุณภาพการเปรียบเทียบโดยใช้กลุ่มสอบทานผล เปรียบเทียบคะแนนด้วยวิธีกำหนดค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าความสามารถผู้สอบให้เท่ากันก่อนการวิเคราะห์ แบบแผนการเก็บรวบรวมข้อมูลใช้กลุ่มสมมูล โมเดล 1 พารามิเตอร์ จำนวนผู้สอบ 700 คน แบบสอบจำนวน 40 ข้อ ได้ดัชนี AMD เท่ากับ 0.000374 ดัชนี MAD เท่ากับ 0.002510 และดัชนี RMS เท่ากับ 0.003022

จากการเปรียบเทียบคะแนนที่หลากหลายเงื่อนไขทำให้ได้ค่าที่เป็นจุดตัดแบ่งดัชนีที่แสดงคุณภาพการเปรียบเทียบคะแนนหลายค่า ดังนั้นจึงได้เฉลี่ยค่าดัชนีและจัดเป็นจุดตัดตามเงื่อนไขการเปรียบเทียบคะแนนดังต่อไปนี้

ตารางที่ 29 ดัชนีความแตกต่าง AMD, MAD และ RMS ที่กำหนดเป็นจุดตัดแบ่งกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับสูงออกเป็นสองกลุ่มย่อย จำแนกตามเงื่อนไขการเปรียบเทียบคะแนน

เงื่อนไขการเปรียบเทียบคะแนน	ค่าดัชนีความแตกต่าง		
	AMD	MAD	RMS
รวมเงื่อนไขทั้งหมด	0.000183	0.002391	0.002907
โมเดล 1 พารามิเตอร์	0.000182	0.002450	0.002924
โมเดล 3 พารามิเตอร์	0.000183	0.002362	0.002899
แบบแผนกลุ่มสมมูล	0.000197	0.002353	0.002924
แบบแผนใช้ข้อสอบร่วม	0.000176	0.002410	0.002899
การเปรียบเทียบกลับสู่แบบสอบเดิม	0.000169	0.002370	0.002875
การใช้กลุ่มสอบทานผล	0.000189	0.002402	0.002924

จากตารางที่ 29 พบว่าค่าดัชนีความแตกต่างที่กำหนดเป็นจุดตัดแบ่งกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับสูงออกเป็นสองกลุ่มย่อยแตกต่างกันตามเงื่อนไขการเปรียบเทียบคะแนน โดยดัชนี AMD เงื่อนไขแบบแผนกลุ่มสมมูลให้ค่าสูงสุด เท่ากับ 0.000197 รองลงมาคือเงื่อนไขการใช้กลุ่มสอบทานผล รวมเงื่อนไขทั้งหมดซึ่งเท่ากับโมเดล 3 พารามิเตอร์ โมเดล 1 พารามิเตอร์ แบบแผนใช้ข้อสอบร่วม และการปรับเทียบกลับสู่แบบสอบเดิม ซึ่งให้ค่าดัชนีเท่ากับ 0.000189, 0.000183, 0.000182, 0.000176 และ 0.000169 ตามลำดับ

ดัชนี MAD เงื่อนไขโมเดล 1 พารามิเตอร์ให้ค่าสูงสุด เท่ากับ 0.002450 รองลงมาคือเงื่อนไขใช้ข้อสอบร่วม การใช้กลุ่มสอบทานผล รวมเงื่อนไขทั้งหมด การปรับกลับสู่แบบสอบเดิม โมเดล 3 พารามิเตอร์ และ การใช้กลุ่มสมมูล ซึ่งให้ค่าดัชนีเท่ากับ 0.002410, 0.002402, 0.002391, 0.002370, 0.002362 และ 0.002353 ตามลำดับ

ดัชนี RMS เงื่อนไขโมเดล 1 พารามิเตอร์ การใช้กลุ่มสมมูล และการใช้กลุ่มสอบทานผลให้ค่าดัชนีสูงสุดเท่ากัน ค่าดัชนีเท่ากับ 0.002924 รองลงมาคือรวมเงื่อนไขทั้งหมด โมเดล 3 พารามิเตอร์ซึ่งให้ค่าดัชนีเท่ากันกับเงื่อนไขการใช้ข้อสอบร่วม และการปรับเทียบกลับสู่แบบสอบเดิม ซึ่งให้ค่าดัชนีเท่ากับ 0.002907, 0.002899 และ 0.002875 ตามลำดับ

ตารางที่ 30 ดัชนีความแตกต่าง AMD, MAD และ RMS ที่กำหนดเป็นจุดตัดแบ่งกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับต่ำออกเป็นสองกลุ่มย่อย จำแนกตามเงื่อนไขการเปรียบเทียบคะแนน

เงื่อนไขการเปรียบเทียบคะแนน	ค่าดัชนีความแตกต่าง		
	AMD	MAD	RMS
รวมเงื่อนไขทั้งหมด	0.000352	0.002621	0.003181
โมเดล 1 พารามิเตอร์	0.000355	0.002574	0.003073
โมเดล 3 พารามิเตอร์	0.000347	0.002717	0.003398
แบบแผนกลุ่มสมมูล	0.000341	0.002677	0.003261
แบบแผนใช้ข้อสอบร่วม	0.000374	0.002510	0.003022
การปรับเทียบกลับสู่แบบสอบเดิม	0.000336	0.002637	0.003124
การใช้กลุ่มสอบทานผล	0.000361	0.002613	0.003210

จากตารางที่ 30 พบว่าค่าดัชนีความแตกต่างที่กำหนดเป็นจุดตัดแบ่งกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับต่ำออกเป็นสองกลุ่มย่อยแตกต่างกันตามเงื่อนไขการเปรียบเทียบคะแนน โดยดัชนี AMD เงื่อนไขแบบแผนการใช้ข้อสอบร่วมให้ค่าสูงสุด เท่ากับ 0.000374

รองลงมาคือเงื่อนไขการใช้กลุ่มสอบทานผล โมเดล 1 พารามิเตอร์ รวมเงื่อนไขทั้งหมด โมเดล 3 พารามิเตอร์ การใช้กลุ่มสมมูล และการปรับเทียบกลับสู่แบบสอบเดิม ซึ่งให้ค่าดัชนีเท่ากับ 0.000361, 0.000355, 0.000352, 0.000347, 0.000341 และ 0.000336 ตามลำดับ

ดัชนี MAD เงื่อนไขโมเดล 3 พารามิเตอร์ให้ค่าสูงสุด เท่ากับ 0.002717 รองลงมาคือเงื่อนไขการใช้กลุ่มสมมูล การปรับเทียบกลับสู่แบบสอบเดิม รวมเงื่อนไขทั้งหมด การใช้กลุ่มสอบทานผล โมเดล 1 พารามิเตอร์ และการใช้ข้อสอบร่วม ซึ่งให้ค่าดัชนีเท่ากับ 0.002677, 0.002637, 0.002621, 0.002613, 0.002574 และ 0.002510 ตามลำดับ

ดัชนี RMS เงื่อนไขโมเดล 3 พารามิเตอร์ ให้ค่าสูงสุด ค่าดัชนีเท่ากับ 0.003398 รองลงมาคือการใช้กลุ่มสมมูล การใช้กลุ่มสอบทานผล การปรับกลับสู่แบบสอบเดิม โมเดล 1 พารามิเตอร์ และการใช้ข้อสอบร่วม ซึ่งให้ค่าดัชนีเท่ากับ 0.003261, 0.003210, 0.003181, 0.003124, 0.003073 และ 0.003022 ตามลำดับ

### 3.2 เกณฑ์ตัดสินคุณภาพการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ

เกณฑ์ตัดสินคุณภาพการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ เป็นกลุ่มของค่าดัชนีความแตกต่างที่บอกระดับคุณภาพการปรับเทียบคะแนน 4 ระดับ คือ ระดับน่าพอใจอย่างยิ่ง ระดับน่าพอใจ ระดับไม่น่าพอใจ และระดับไม่น่าพอใจอย่างยิ่ง ที่กำหนดจากจุดตัด 3 จุด ซึ่งจุดตัดนี้เป็นค่าดัชนีความแตกต่างที่ทดสอบความแตกต่างระหว่างค่าเฉลี่ยของคะแนนจริงที่ปรับเทียบแล้วกับคะแนนจริงที่ยังไม่ได้ปรับเทียบ สำหรับผู้สอบกลุ่มเดียวกันด้วยสถิติ t-test แล้วพบค่าวิกฤต จำแนกเกณฑ์ตามสถานการณ์การปรับเทียบคะแนนเป็น 7 เกณฑ์ดังนี้

#### 3.2.1 เกณฑ์ตัดสินคุณภาพการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ เกณฑ์รวมทุกเงื่อนไขการปรับเทียบคะแนน

ตารางที่ 31 จุดตัดที่เป็นค่าดัชนีความแตกต่างระหว่างคะแนนที่ปรับเทียบแล้วและยังไม่ได้ปรับเทียบ สำหรับเกณฑ์รวมเงื่อนไขทั้งหมดของการปรับเทียบ

จุดตัด	ดัชนี AMD	ดัชนี MAD	ดัชนี RMS
จุดตัดที่แบ่งกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงออกเป็น 2 กลุ่มย่อย	0.000183	0.002391	0.002907
จุดตัดที่แบ่งกลุ่มดัชนีออกเป็นกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงกับระดับต่ำ	0.000244	0.002481	0.003032
จุดตัดที่แบ่งกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำออกเป็น 2 กลุ่มย่อย	0.000352	0.002621	0.003181

จากตารางที่ 31 พบว่าจุดตัดที่แบ่งกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับต่ำ ออกเป็น 2 กลุ่มย่อย ให้ค่าดัชนีความแตกต่างสูงสุด ค่าดัชนี AMD เท่ากับ 0.000352 ค่าดัชนี MAD เท่ากับ 0.002621 และค่าดัชนี RMS เท่ากับ 0.003181 จุดตัดที่ให้ค่าดัชนีความแตกต่างรองลงมา คือจุดตัดที่แบ่งกลุ่มดัชนีออกเป็นกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับสูง กับระดับต่ำ และจุดตัดที่แบ่งกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับสูงออกเป็น 2 กลุ่มย่อย ตามลำดับ

จุดตัดที่ได้จากตารางที่ 31 จะแบ่งกลุ่มดัชนีความแตกต่างออกเป็น 4 กลุ่ม นำเอาดัชนีทั้ง 4 กลุ่มไปจัดเป็นเกณฑ์ตัดสินคุณภาพการเปรียบเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ เป็นเกณฑ์รวมเงื่อนไขทั้งหมด ดังรายละเอียดต่อไปนี้

เมื่อกำหนด

$$AMD = \frac{\left| \frac{\sum_{n=1}^N (\tau - \tau^*)}{N} \right|}{SD}$$

$$MAD = \frac{\frac{\sum_{n=1}^N |\tau - \tau^*|}{N}}{SD}$$

$$RMS = \frac{\sqrt{\frac{\sum_{n=1}^N (\tau - \tau^*)^2}{N}}}{SD}$$

และ  $\tau$  คือ คะแนนจริงจากแบบสอบฉบับที่ 1 สำหรับการเปรียบเทียบกลับสู่แบบสอบเดิม และเป็นคะแนนจริงแบบสอบฉบับที่ 2 สำหรับการเปรียบเทียบใช้กลุ่มสอบทานผล

$\tau^*$  คือ คะแนนจริงจากแบบสอบฉบับที่ 1 ที่ปรับสเกลผ่านแบบสอบฉบับที่ 2 และฉบับที่ 3 เข้าสู่แบบสอบฉบับที่ 1 ในการเปรียบเทียบกลับสู่แบบสอบเดิม และเป็นคะแนนจริงจากฉบับที่ 1 ที่ปรับสเกลเข้าสู่แบบสอบชุดที่ 2 สำหรับกลุ่มสอบทานผล

$N$  คือ จำนวนผู้สอบ

$SD$  คือ ส่วนเบี่ยงเบนมาตรฐานเฉลี่ยจากคะแนน  $\tau$  และ คะแนน  $\tau^*$



ตารางที่ 32 เกณฑ์ตัดสินคุณภาพการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ  
เกณฑ์รวมเงื่อนไขทั้งหมด

ระดับคุณภาพการปรับเทียบ	ดัชนีความแตกต่าง
น่าพอใจอย่างยิ่ง	AMD < 0.000183
	MAD < 0.002391
	RMS < 0.002907
น่าพอใจ	$0.000183 \leq \text{AMD} < 0.000244$
	$0.002391 \leq \text{MAD} < 0.002481$
	$0.002907 \leq \text{RMS} < 0.003032$
ไม่น่าพอใจ	$0.000244 \leq \text{AMD} < 0.000352$
	$0.002481 \leq \text{MAD} < 0.002621$
	$0.003032 \leq \text{RMS} < 0.003181$
ไม่น่าพอใจอย่างยิ่ง	$0.000352 \leq \text{AMD}$
	$0.002621 \leq \text{MAD}$
	$0.003181 \leq \text{RMS}$

### 3.2.2 เกณฑ์ตัดสินคุณภาพการปรับเทียบคะแนน เกณฑ์สำหรับโมเดล 1 พารามิเตอร์ ตามทฤษฎีการตอบสนองข้อสอบ

ตารางที่ 33 จุดตัดที่เป็นค่าดัชนีความแตกต่างระหว่างคะแนนที่ปรับเทียบแล้วและยังไม่ได้  
ปรับเทียบ สำหรับเกณฑ์โมเดล 1 พารามิเตอร์ ตามทฤษฎีการตอบสนองข้อสอบ

จุดตัด	ดัชนี AMD	ดัชนี MAD	ดัชนี RMS
จุดตัดที่แบ่งกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบ ระดับสูงออกเป็น 2 กลุ่มย่อย	0.000182	0.002450	0.002924
จุดตัดที่แบ่งกลุ่มดัชนีออกเป็นกลุ่มดัชนีที่แสดง คุณภาพการปรับเทียบระดับสูงกับระดับต่ำ	0.000261	0.002508	0.003045
จุดตัดที่แบ่งกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบ ระดับต่ำออกเป็น 2 กลุ่มย่อย	0.000355	0.002574	0.003073

จากตารางที่ 33 พบว่าจุดตัดที่แบ่งกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำ ออกเป็น 2 กลุ่มย่อย ให้ค่าดัชนีความแตกต่างสูงสุด ค่าดัชนี AMD เท่ากับ 0.000355 ค่าดัชนี MAD เท่ากับ 0.002574 และค่าดัชนี RMS เท่ากับ 0.003073 จุดตัดที่ให้ค่าดัชนีความแตกต่างรองลงมา คือจุดตัดที่แบ่งกลุ่มดัชนีออกเป็นกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูง กับระดับต่ำ และจุดตัดที่แบ่งกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงออกเป็น 2 กลุ่มย่อย ตามลำดับ

จุดตัดที่ได้จากตารางที่ 33 จะแบ่งกลุ่มดัชนีความแตกต่างออกเป็น 4 กลุ่ม นำเอาดัชนีทั้ง 4 กลุ่มไปจัดเป็นเกณฑ์ตัดสินคุณภาพการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ เป็นเกณฑ์สำหรับโมเดล 1 พารามิเตอร์ ดังตารางที่ 34

ตารางที่ 34 เกณฑ์ตัดสินคุณภาพการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ  
เกณฑ์สำหรับโมเดล 1 พารามิเตอร์

ระดับคุณภาพการปรับเทียบ	ดัชนีความแตกต่าง
น่าพอใจอย่างยิ่ง	AMD < 0.000182
	MAD < 0.002450
	RMS < 0.002924
น่าพอใจ	$0.000182 \leq \text{AMD} < 0.000261$
	$0.002450 \leq \text{MAD} < 0.002508$
	$0.002924 \leq \text{RMS} < 0.003045$
ไม่น่าพอใจ	$0.000261 \leq \text{AMD} < 0.000355$
	$0.002508 \leq \text{MAD} < 0.002574$
	$0.003045 \leq \text{RMS} < 0.003073$
ไม่น่าพอใจอย่างยิ่ง	$0.000355 \leq \text{AMD}$
	$0.002574 \leq \text{MAD}$
	$0.003073 \leq \text{RMS}$

### 3.2.3 เกณฑ์ตัดสินคุณภาพการปรับเทียบคะแนน เกณฑ์สำหรับโมเดล 3 พารามิเตอร์ ตามทฤษฎีการตอบสนองข้อสอบ

ตารางที่ 35 จุดตัดที่เป็นค่าดัชนีความแตกต่างระหว่างคะแนนที่ปรับเทียบแล้วและยังไม่ได้ปรับเทียบ สำหรับเกณฑ์โมเดล 3 พารามิเตอร์ ตามทฤษฎีการตอบสนองข้อสอบ

จุดตัด	ดัชนี AMD	ดัชนี MAD	ดัชนี RMS
จุดตัดที่แบ่งกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงออกเป็น 2 กลุ่มย่อย	0.000183	0.002402	0.002924
จุดตัดที่แบ่งกลุ่มดัชนีออกเป็นกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงกับระดับต่ำ	0.000242	0.002468	0.003122
จุดตัดที่แบ่งกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำออกเป็น 2 กลุ่มย่อย	0.000355	0.002613	0.003398

จากตารางที่ 35 พบว่าจุดตัดที่แบ่งกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำออกเป็น 2 กลุ่มย่อย ให้ค่าดัชนีความแตกต่างสูงสุด ค่าดัชนี AMD เท่ากับ 0.000355 ค่าดัชนี MAD เท่ากับ 0.002613 และค่าดัชนี RMS เท่ากับ 0.003398 จุดตัดที่ให้ค่าดัชนีความแตกต่างรองลงมา คือจุดตัดที่แบ่งกลุ่มดัชนีออกเป็นกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงกับระดับต่ำ และจุดตัดที่แบ่งกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงออกเป็น 2 กลุ่มย่อย ตามลำดับ

จุดตัดที่ได้จากตารางที่ 35 จะแบ่งกลุ่มดัชนีความแตกต่างออกเป็น 4 กลุ่ม นำเอาดัชนีทั้ง 4 กลุ่มไปจัดเป็นเกณฑ์ตัดสินคุณภาพการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ เป็นเกณฑ์สำหรับโมเดล 3 พารามิเตอร์ ดังตารางที่ 36

ตารางที่ 36 เกณฑ์ตัดสินคุณภาพการเปรียบเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ  
เกณฑ์สำหรับโมเดล 3 พารามิเตอร์

ระดับคุณภาพการเปรียบเทียบ	ดัชนีความแตกต่าง
นำพอใจอย่างยิ่ง	AMD < 0.000183
	MAD < 0.002402
	RMS < 0.002924
นำพอใจ	$0.000183 \leq \text{AMD} < 0.000242$
	$0.002402 \leq \text{MAD} < 0.002468$
	$0.002924 \leq \text{RMS} < 0.003122$
ไม่นำพอใจ	$0.000242 \leq \text{AMD} < 0.000355$
	$0.002468 \leq \text{MAD} < 0.002613$
	$0.003122 \leq \text{RMS} < 0.003398$
ไม่นำพอใจอย่างยิ่ง	$0.000355 \leq \text{AMD}$
	$0.002613 \leq \text{MAD}$
	$0.003398 \leq \text{RMS}$

### 3.2.4 เกณฑ์ตัดสินคุณภาพการเปรียบเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ สำหรับแบบแผนการเก็บรวบรวมข้อมูลใช้กลุ่มสมมูล

ตารางที่ 37 จุดตัดที่เป็นค่าดัชนีความแตกต่างระหว่างคะแนนที่เปรียบเทียบแล้วและยังไม่ได้  
เปรียบเทียบ สำหรับเกณฑ์แบบแผนการเก็บรวบรวมข้อมูลใช้กลุ่มสมมูล

จุดตัด	ดัชนี AMD	ดัชนี MAD	ดัชนี RMS
จุดตัดที่แบ่งกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับสูงออกเป็น 2 กลุ่มย่อย	0.000197	0.002353	0.002924
จุดตัดที่แบ่งกลุ่มดัชนีออกเป็นกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับสูงกับระดับต่ำ	0.000237	0.002375	0.002930
จุดตัดที่แบ่งกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับต่ำออกเป็น 2 กลุ่มย่อย	0.000374	0.002510	0.003022

จากตารางที่ 37 พบว่าจุดตัดที่แบ่งกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับต่ำ ออกเป็น 2 กลุ่มย่อย ให้ค่าดัชนีความแตกต่างสูงสุด ค่าดัชนี AMD เท่ากับ 0.000374 ค่าดัชนี MAD เท่ากับ 0.002510 และค่าดัชนี RMS เท่ากับ 0.003022 จุดตัดที่ให้ค่าดัชนีความแตกต่างรองลงมา คือจุดตัดที่แบ่งกลุ่มดัชนีออกเป็นกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับสูง กับระดับต่ำ และจุดตัดที่แบ่งกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับสูงออกเป็น 2 กลุ่มย่อย ตามลำดับ

จุดตัดที่ได้จากตารางที่ 37 จะแบ่งกลุ่มดัชนีความแตกต่างออกเป็น 4 กลุ่ม นำเอาดัชนีทั้ง 4 กลุ่มไปจัดเป็นเกณฑ์ตัดสินคุณภาพการเปรียบเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ เป็นเกณฑ์สำหรับการเก็บรวบรวมข้อมูลใช้กลุ่มสมมูล ดังตารางที่ 38

ตารางที่ 38 เกณฑ์ตัดสินคุณภาพการเปรียบเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ  
เกณฑ์สำหรับแบบแผนการเก็บรวบรวมข้อมูลใช้กลุ่มสมมูล

ระดับคุณภาพการเปรียบเทียบ	ดัชนีความแตกต่าง
น่าพอใจอย่างยิ่ง	AMD < 0.000197
	MAD < 0.002353
	RMS < 0.002924
น่าพอใจ	$0.000197 \leq \text{AMD} < 0.000237$
	$0.002353 \leq \text{MAD} < 0.002375$
	$0.002924 \leq \text{RMS} < 0.002930$
ไม่น่าพอใจ	$0.000237 \leq \text{AMD} < 0.000374$
	$0.002375 \leq \text{MAD} < 0.002510$
	$0.002930 \leq \text{RMS} < 0.003022$
ไม่น่าพอใจอย่างยิ่ง	$0.000374 \leq \text{AMD}$
	$0.002510 \leq \text{MAD}$
	$0.003022 \leq \text{RMS}$

### 3.2.5 เกณฑ์ตัดสินคุณภาพการเปรียบเทียบคะแนนตามทฤษฎีการตอบสนอง ข้อสอบ สำหรับรูปแบบการเก็บรวบรวมข้อมูลใช้ข้อสอบร่วม

ตารางที่ 39 จุดตัดที่เป็นค่าดัชนีความแตกต่างระหว่างคะแนนที่เปรียบเทียบแล้วและยังไม่ได้  
เปรียบเทียบ สำหรับเกณฑ์แบบแผนการเก็บรวบรวมข้อมูลใช้ข้อสอบร่วม

จุดตัด	ดัชนี AMD	ดัชนี MAD	ดัชนี RMS
จุดตัดที่แบ่งกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบ ระดับสูงออกเป็น 2 กลุ่มย่อย	0.000176	0.002410	0.002899
จุดตัดที่แบ่งกลุ่มดัชนีออกเป็นกลุ่มดัชนีที่แสดง คุณภาพการเปรียบเทียบระดับสูงกับระดับต่ำ	0.000247	0.002535	0.003083
จุดตัดที่แบ่งกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบ ระดับต่ำออกเป็น 2 กลุ่มย่อย	0.000341	0.002677	0.003261

จากตารางที่ 39 พบว่าจุดตัดที่แบ่งกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับต่ำ  
ออกเป็น 2 กลุ่มย่อย ให้ค่าดัชนีความแตกต่างสูงสุด ค่าดัชนี AMD เท่ากับ 0.000341 ค่าดัชนี  
MAD เท่ากับ 0.002677 และค่าดัชนี RMS เท่ากับ 0.003261 จุดตัดที่ให้ค่าดัชนีความแตก  
ต่างรองลงมา คือจุดตัดที่แบ่งกลุ่มดัชนีออกเป็นกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับสูง  
กับระดับต่ำ และจุดตัดที่แบ่งกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับสูงออกเป็น 2 กลุ่ม  
ย่อย ตามลำดับ

จุดตัดที่ได้จากตารางที่ 39 จะแบ่งกลุ่มดัชนีความแตกต่างออกเป็น 4 กลุ่ม นำเอา  
ดัชนีทั้ง 4 กลุ่มไปจัดเป็นเกณฑ์ตัดสินคุณภาพการเปรียบเทียบคะแนนตามทฤษฎีการตอบสนอง  
ข้อสอบ เป็นเกณฑ์สำหรับการเก็บรวบรวมข้อมูลใช้ข้อสอบร่วม ดังตารางที่ 40

ตารางที่ 40 เกณฑ์ตัดสินคุณภาพการเปรียบเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ  
เกณฑ์สำหรับแบบแผนการเก็บรวบรวมข้อมูลใช้ข้อสอบร่วม

ระดับคุณภาพการเปรียบเทียบ	ดัชนีความแตกต่าง
น่าพอใจอย่างยิ่ง	AMD < 0.000176
	MAD < 0.002410
	RMS < 0.002899
น่าพอใจ	$0.000176 \leq \text{AMD} < 0.000247$
	$0.002410 \leq \text{MAD} < 0.002535$
	$0.002899 \leq \text{RMS} < 0.003083$
ไม่น่าพอใจ	$0.000247 \leq \text{AMD} < 0.000341$
	$0.002535 \leq \text{MAD} < 0.002677$
	$0.003083 \leq \text{RMS} < 0.003261$
ไม่น่าพอใจอย่างยิ่ง	$0.000341 \leq \text{AMD}$
	$0.002677 \leq \text{MAD}$
	$0.003261 \leq \text{RMS}$

3.2.6 เกณฑ์ตัดสินคุณภาพการเปรียบเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ สำหรับการหาคุณภาพการเปรียบเทียบคะแนนด้วยการเปรียบเทียบคะแนนกลับสู่แบบสอบเดิม

ตารางที่ 41 จุดตัดที่เป็นค่าดัชนีความแตกต่างระหว่างคะแนนที่เปรียบเทียบแล้วและยังไม่ได้เปรียบเทียบ สำหรับการหาคุณภาพการเปรียบเทียบด้วยการเปรียบเทียบกลับสู่แบบสอบเดิม

จุดตัด	ดัชนี AMD	ดัชนี MAD	ดัชนี RMS
จุดตัดที่แบ่งกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับสูงออกเป็น 2 กลุ่มย่อย	0.000169	0.002370	0.002875
จุดตัดที่แบ่งกลุ่มดัชนีออกเป็นกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับสูงกับระดับต่ำ	0.000261	0.002508	0.003045
จุดตัดที่แบ่งกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับต่ำออกเป็น 2 กลุ่มย่อย	0.000336	0.002637	0.003124

จากตารางที่ 41 พบว่าจุดตัดที่แบ่งกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำ ออกเป็น 2 กลุ่มย่อย ให้ค่าดัชนีความแตกต่างสูงสุด ค่าดัชนี AMD เท่ากับ 0.000336 ค่าดัชนี MAD เท่ากับ 0.002637 และค่าดัชนี RMS เท่ากับ 0.003124 จุดตัดที่ให้ค่าดัชนีความแตกต่างรองลงมา คือจุดตัดที่แบ่งกลุ่มดัชนีออกเป็นกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูง กับระดับต่ำ และจุดตัดที่แบ่งกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงออกเป็น 2 กลุ่มย่อย ตามลำดับ

จุดตัดที่ได้จากตารางที่ 41 จะแบ่งกลุ่มดัชนีความแตกต่างออกเป็น 4 กลุ่ม นำเอาดัชนีทั้ง 4 กลุ่มไปจัดเป็นเกณฑ์ตัดสินคุณภาพการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ เป็นเกณฑ์สำหรับการปรับเทียบกลับสู่แบบสอบเดิม ดังตารางที่ 42

ตารางที่ 42 เกณฑ์ตัดสินคุณภาพการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ  
เกณฑ์สำหรับการปรับเทียบกลับสู่แบบสอบเดิม

ระดับคุณภาพการปรับเทียบ	ดัชนีความแตกต่าง
นำพอใจอย่างยิ่ง	AMD < 0.000169
	MAD < 0.002370
	RMS < 0.002875
นำพอใจ	$0.000169 \leq \text{AMD} < 0.000261$
	$0.002370 \leq \text{MAD} < 0.002508$
	$0.002875 \leq \text{RMS} < 0.003045$
ไม่นำพอใจ	$0.000261 \leq \text{AMD} < 0.000336$
	$0.002508 \leq \text{MAD} < 0.002637$
	$0.003045 \leq \text{RMS} < 0.003124$
ไม่นำพอใจอย่างยิ่ง	$0.000336 \leq \text{AMD}$
	$0.002637 \leq \text{MAD}$
	$0.003124 \leq \text{RMS}$



### 3.2.7 เกณฑ์ตัดสินคุณภาพการเปรียบเทียบคะแนนตามทฤษฎีการตอบสนอง ข้อสอบ สำหรับการหาคุณภาพการเปรียบเทียบคะแนนด้วยการใช้กลุ่มสอบทานผล

ตารางที่ 43 จุดตัดที่เป็นค่าดัชนีความแตกต่างระหว่างคะแนนที่เปรียบเทียบแล้วและยังไม่ได้  
เปรียบเทียบ สำหรับเกณฑ์การหาคุณภาพการเปรียบเทียบด้วยการใช้กลุ่มสอบทานผล

จุดตัด	ดัชนี AMD	ดัชนี MAD	ดัชนี RMS
จุดตัดที่แบ่งกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบ ระดับสูงออกเป็น 2 กลุ่มย่อย	0.000189	0.002402	0.002924
จุดตัดที่แบ่งกลุ่มดัชนีออกเป็นกลุ่มดัชนีที่แสดง คุณภาพการเปรียบเทียบระดับสูงกับระดับต่ำ	0.000235	0.002468	0.003026
จุดตัดที่แบ่งกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบ ระดับต่ำออกเป็น 2 กลุ่มย่อย	0.000361	0.002613	0.003210

จากตารางที่ 43 พบว่าจุดตัดที่แบ่งกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับต่ำ  
ออกเป็น 2 กลุ่มย่อย ให้ค่าดัชนีความแตกต่างสูงสุด ค่าดัชนี AMD เท่ากับ 0.000361 ค่าดัชนี  
MAD เท่ากับ 0.002613 และค่าดัชนี RMS เท่ากับ 0.003210 จุดตัดที่ให้ค่าดัชนีความแตก  
ต่างรองลงมา คือจุดตัดที่แบ่งกลุ่มดัชนีออกเป็นกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับสูง  
กับระดับต่ำ และจุดตัดที่แบ่งกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับสูงออกเป็น 2 กลุ่ม  
ย่อย ตามลำดับ

จุดตัดที่ได้จากตารางที่ 43 จะแบ่งกลุ่มดัชนีความแตกต่างออกเป็น 4 กลุ่ม นำเอา  
ดัชนีทั้ง 4 กลุ่มไปจัดเป็นเกณฑ์ตัดสินคุณภาพการเปรียบเทียบคะแนนตามทฤษฎีการตอบสนอง  
ข้อสอบ เป็นเกณฑ์สำหรับการหาคุณภาพการเปรียบเทียบคะแนนด้วยการใช้กลุ่มสอบทานผล  
ดังตารางที่ 44



ตารางที่ 44 เกณฑ์ตัดสินคุณภาพการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ  
เกณฑ์สำหรับการปรับเทียบกลับสู่แบบสอบเดิม

ระดับคุณภาพการปรับเทียบ	ดัชนีความแตกต่าง
น่าพอใจอย่างยิ่ง	AMD < 0.000189
	MAD < 0.002402
	RMS < 0.002924
น่าพอใจ	$0.000189 \leq \text{AMD} < 0.000235$
	$0.002402 \leq \text{MAD} < 0.002468$
	$0.002924 \leq \text{RMS} < 0.003026$
ไม่น่าพอใจ	$0.000235 \leq \text{AMD} < 0.000361$
	$0.002468 \leq \text{MAD} < 0.002613$
	$0.003026 \leq \text{RMS} < 0.003210$
ไม่น่าพอใจอย่างยิ่ง	$0.000361 \leq \text{AMD}$
	$0.002613 \leq \text{MAD}$
	$0.003210 \leq \text{RMS}$

#### ตอนที่ 4 ผลการตรวจสอบคุณภาพของเกณฑ์ที่พัฒนาขึ้น

การวิจัยครั้งนี้ได้ตรวจสอบคุณภาพของเกณฑ์ที่พัฒนาขึ้น เป็นการหาความสอดคล้องกับเกณฑ์ความเสมอภาคของลอร์ดและเกณฑ์ของปีเตอร์เซนและคณะ โดยได้จำลองข้อมูลที่มีผลการปรับเทียบคะแนนในระดับคุณภาพน่าพอใจอย่างยิ่งในเงื่อนไขโมเดล 3 พารามิเตอร์ แบบแผนการเก็บรวบรวมข้อมูลใช้ข้อสอบร่วม หากคุณภาพการปรับเทียบด้วยการใช้กลุ่มสอบทานผลปรับเทียบคะแนนด้วยวิธี Mean and Sigma ข้อสอบ 90 ข้อ ผู้สอบ 1,200 คน ระดับคุณภาพการปรับเทียบน่าพอใจในเงื่อนไขโมเดล 1 พารามิเตอร์ แบบแผนใช้ข้อสอบร่วม ใช้กลุ่มสอบทานผล ปรับเทียบคะแนนโดยวิธี Mean and Sigma ข้อสอบ 60 ข้อ ผู้สอบ 1,100 คน ระดับคุณภาพการปรับเทียบไม่น่าพอใจในเงื่อนไขโมเดล 3 พารามิเตอร์ แบบแผนกลุ่มสมมูล ใช้กลุ่มสอบทานผล ปรับเทียบคะแนนด้วยวิธี Same Scaling Convention ข้อสอบ 80 ข้อ ผู้สอบ 1,300 คน และคุณภาพการปรับเทียบไม่น่าพอใจอย่างยิ่งในเงื่อนไขโมเดล 3 พารามิเตอร์ แบบแผนใช้ข้อสอบร่วม ใช้การปรับกลับสู่แบบสอบเดิม ปรับเทียบคะแนนด้วยวิธี Mean and Mean ข้อสอบ 70 ข้อ และผู้สอบ 1,400 คน ผลการวิเคราะห์มีดังต่อไปนี้

ตารางที่ 45 การทดสอบความแตกต่างของค่าเฉลี่ยคะแนนจริงที่ไม่ได้ปรับกับค่าเฉลี่ยคะแนนจริงที่ปรับเทียบแล้ว จำแนกตามระดับคุณภาพการปรับเทียบจากข้อมูลที่จำลองขึ้นตามเงื่อนไขการปรับเทียบคะแนน ใช้เกณฑ์ที่พัฒนาขึ้น

ระดับคุณภาพการปรับเทียบ	ค่าเฉลี่ยคะแนนจริง		S.D. ของคะแนนจริง		ค่าความต่างเฉลี่ย	t	$\alpha$
	ไม่ปรับ	ปรับแล้ว	ไม่ปรับ	ปรับแล้ว			
นำพอใจอย่างยิ่ง	58.787	58.786	17.440	17.438	0.00072	0.503	.615
นำพอใจ	30.490	30.485	13.544	13.545	0.00437	2.477	.013
ไม่นำพอใจ	49.298	49.294	14.940	14.938	0.00398	3.042	.002
ไม่นำพอใจอย่างยิ่ง	44.277	46.708	13.284	12.144	2.43101	66.51	.000

จากตารางที่ 45 พบว่าการปรับเทียบคุณภาพไม่นำพอใจอย่างยิ่งให้ค่าความต่างเฉลี่ยมากที่สุด เท่ากับ 2.43101 รองลงมาคือการปรับเทียบคุณภาพไม่นำพอใจ คุณภาพนำพอใจและคุณภาพนำพอใจอย่างยิ่งตามลำดับ ส่วนค่า  $\alpha$  การปรับเทียบคุณภาพนำพอใจอย่างยิ่งให้ค่าสูงสุดเท่ากับ .615 รองลงมาคือการปรับเทียบคุณภาพนำพอใจ คุณภาพไม่นำพอใจ และคุณภาพนำพอใจอย่างยิ่งตามลำดับ

นำคะแนนจริงที่ไม่ได้ปรับเทียบและคะแนนจริงที่ปรับเทียบแล้วจากเงื่อนไขการปรับเทียบคะแนน ไปคำนวณค่าดัชนีความแตกต่าง AMD, MAD, และ RMS ได้ผลดังตารางต่อไปนี้

ตารางที่ 46 ค่าดัชนีความแตกต่าง AMD, MAD, และ RMS จำแนกตามระดับคุณภาพการปรับเทียบตามเงื่อนไขการปรับเทียบคะแนน

ระดับคุณภาพการปรับเทียบ	ค่าดัชนี		
	ดัชนี AMD	ดัชนี MAD	ดัชนี RMS
นำพอใจอย่างยิ่ง	0.000042	0.002340	0.002858
นำพอใจ	0.000196	0.002380	0.002889
ไม่นำพอใจ	0.000283	0.002554	0.003067
ไม่นำพอใจอย่างยิ่ง	0.191205	0.192479	0.219366

จากตารางที่ 46 พบว่า การเปรียบเทียบคุณภาพไม่นำพ้อใจอย่างยิ่งให้ทั้งค่าดัชนี AMD, MAD และ RMS สูงสุด ค่าดัชนีเท่ากับ 0.191205, 0.192479 และ 0.219366 ตามลำดับ รองลงมา คือการเปรียบเทียบคะแนนคุณภาพไม่นำพ้อใจ คุณภาพนำพ้อใจ และคุณภาพนำพ้อใจอย่างยิ่ง ตามลำดับ

นำข้อมูลที่จำลองขึ้นตามเงื่อนไขการเปรียบเทียบคะแนนที่มี 4 ระดับคุณภาพการเปรียบเทียบคะแนน ไปวิเคราะห์หาค่าดัชนีความแตกต่างตามเกณฑ์ของปีเตอร์เซนและคณะ ได้ผลดังต่อไปนี้

ตารางที่ 47 ค่าดัชนีความแตกต่างตามเกณฑ์ของปีเตอร์เซนและคณะ

ระดับคุณภาพของเกณฑ์ที่พัฒนาขึ้น	ค่าดัชนี MSE ตามเกณฑ์ของปีเตอร์เซน	เกณฑ์ของปีเตอร์เซน
นำพ้อใจอย่างยิ่ง	0.0000082	พ้อใจอย่างมาก MSE < 0.760496 นำพ้อใจ 0.760496 ≤ MSE < 3.041982 ปานกลาง 3.041982 ≤ MSE < 6.844460 ไม่นำพ้อใจ 6.844460 ≤ MSE < 12.167929 ไม่นำพ้อใจอย่างยิ่ง 12.167929 ≤ MSE
นำพ้อใจ	0.0000087	พ้อใจอย่างมาก MSE < 0.458725 นำพ้อใจ 0.458725 ≤ MSE < 1.834902 ปานกลาง 1.834902 ≤ MSE < 4.128529 ไม่นำพ้อใจ 4.128529 ≤ MSE < 7.339609 ไม่นำพ้อใจอย่างยิ่ง 7.339609 ≤ MSE
ไม่นำพ้อใจ	0.0000091	พ้อใจอย่างมาก MSE < 0.558027 นำพ้อใจ 0.558027 ≤ MSE < 2.232109 ปานกลาง 2.232109 ≤ MSE < 5.022246 ไม่นำพ้อใจ 5.022246 ≤ MSE < 8.928438 ไม่นำพ้อใจอย่างยิ่ง 8.928438 ≤ MSE
ไม่นำพ้อใจอย่างยิ่ง	0.0440792	พ้อใจอย่างมาก MSE < 0.441184 นำพ้อใจ 0.441184 ≤ MSE < 1.764737 ปานกลาง 1.764737 ≤ MSE < 3.970658 ไม่นำพ้อใจ 3.970658 ≤ MSE < 7.058947 ไม่นำพ้อใจอย่างยิ่ง 7.058947 ≤ MSE

จากตารางที่ 47 พบว่าดัชนีที่แสดงคุณภาพการปรับเทียบคะแนนทั้ง 4 ระดับ คือ ระดับนำพอใจอย่างยิ่ง ระดับนำพอใจ ระดับไม่นำพอใจ และระดับไม่นำพอใจอย่างยิ่ง จากเกณฑ์ที่พัฒนาขึ้น จะอยู่ในระดับคุณภาพการปรับเทียบนำพอใจอย่างยิ่ง เพียงระดับเดียวเมื่อใช้เกณฑ์ของปีเตอร์เซนและคณะ

นำข้อมูลที่จำลองขึ้นตามเงื่อนไขการปรับเทียบคะแนนที่มี 4 ระดับคุณภาพการปรับเทียบคะแนน ไปวิเคราะห์ตามเกณฑ์ความเสมอภาคของลอร์ด โดยการทดสอบความแตกต่างระหว่างการแจกแจงของคะแนนจริงที่ไม่ได้ปรับและการแจกแจงของคะแนนจริงที่ปรับเทียบแล้ว ด้วยสถิติ Wilcoxon sign-rank test ได้ผลดังต่อไปนี้

ตารางที่ 48 ทดสอบความแตกต่างระหว่างการแจกแจงคะแนนจริงที่ปรับเทียบแล้วกับคะแนนจริงที่ยังไม่ปรับเทียบ ด้วยสถิติ Wilcoxon sign-rank test สำหรับการปรับเทียบคะแนนระดับคุณภาพทั้ง 4 ระดับ

ระดับคุณภาพการปรับเทียบ	ค่าสถิติ Z
นำพอใจอย่างยิ่ง	0.317
นำพอใจ	2.255
ไม่นำพอใจ	3.234**
ไม่นำพอใจอย่างยิ่ง	32.034**

\*\* P < .01

จากตารางที่ 48 พบว่า การแจกแจงคะแนนจริงที่ไม่ได้ปรับเทียบและการแจกแจงคะแนนจริงที่ปรับเทียบแล้ว แตกต่างกันอย่างมีนัยสำคัญที่ระดับ .01 ในเงื่อนไขระดับคุณภาพการปรับเทียบไม่นำพอใจอย่างยิ่ง และระดับคุณภาพการปรับเทียบไม่นำพอใจ ส่วนเงื่อนไขการปรับเทียบคุณภาพนำพอใจอย่างยิ่งและระดับคุณภาพนำพอใจ การแจกแจงคะแนนจริงทั้งสองไม่แตกต่างกัน ซึ่งจะบอกถึงผลการปรับเทียบคะแนนในระดับนำพอใจเมื่อใช้เกณฑ์ความเสมอภาคของลอร์ด



จากการหาคุณภาพของเกณฑ์ที่พัฒนาขึ้นด้วยการหาความสอดคล้องกับเกณฑ์ความเสมอภาคของลอร์ด และเกณฑ์ของปีเตอร์เซนและคณะ จำแนกเกณฑ์ตามเงื่อนไขการปรับเทียบคะแนน ดังต่อไปนี้

ตารางที่ 49 ความสอดคล้องของเกณฑ์ที่พัฒนาขึ้นสำหรับเกณฑ์รวมทุกเงื่อนไข กับเกณฑ์ความเสมอภาคของลอร์ด และเกณฑ์ของปีเตอร์เซนและคณะ

เกณฑ์ที่พัฒนาขึ้น	ระดับคุณภาพการปรับเทียบ							
	เกณฑ์ของปีเตอร์เซนและคณะ							
	เกณฑ์ของลอร์ด		เกณฑ์ของปีเตอร์เซนและคณะ		เกณฑ์ของปีเตอร์เซนและคณะ		เกณฑ์ของลอร์ด	
	นำพอใจอย่างมาก	นำพอใจ	ปานกลาง	ไม่นำพอใจ	ไม่นำพอใจอย่างมาก	นำพอใจ	ไม่นำพอใจ	
นำพอใจอย่างยิ่ง	√					√		
นำพอใจ	√					√		
ไม่นำพอใจ	√						√	
ไม่นำพอใจอย่างยิ่ง	√						√	

จากตารางที่ 49 ผลการตัดสินคุณภาพการปรับเทียบคะแนนเมื่อใช้เกณฑ์ที่พัฒนาขึ้นเปรียบเทียบกับผลการใช้ตามเกณฑ์ของปีเตอร์เซนและคณะ พบว่าไม่สอดคล้องกัน โดยผลการตัดสินคุณภาพการปรับเทียบทั้ง 4 ระดับจากเกณฑ์ที่พัฒนาขึ้น อยู่ในระบับนำพอใจอย่างมาก เมื่อเทียบกับเกณฑ์ของปีเตอร์เซนและคณะ ส่วนผลการตัดสินคุณภาพการปรับเทียบคะแนนเมื่อใช้เกณฑ์ที่พัฒนาขึ้นเปรียบเทียบกับผลการใช้เกณฑ์ความเสมอภาคของลอร์ด พบว่ามีความสอดคล้องกัน โดยผลการตัดสินคุณภาพการปรับเทียบระดับนำพอใจอย่างยิ่งหรือระดับนำพอใจ และระดับไม่นำพอใจหรือระดับไม่นำพอใจอย่างยิ่งจากการใช้เกณฑ์ที่พัฒนาขึ้น อยู่ในระดับคุณภาพนำพอใจและระดับคุณภาพไม่นำพอใจตามลำดับ เมื่อใช้เกณฑ์ความเสมอภาคของลอร์ด

ตารางที่ 50 ความสอดคล้องของเกณฑ์ที่พัฒนาขึ้นสำหรับเกณฑ์โมเดล 1 พารามิเตอร์ กับ  
เกณฑ์ความเสมอภาคของลอร์ด และเกณฑ์ของปีเตอร์เซนและคณะ

เกณฑ์ที่พัฒนาขึ้น	ระดับคุณภาพการเปรียบเทียบ						
	เกณฑ์ของปีเตอร์เซนและคณะ					เกณฑ์ของลอร์ด	
	นำพอใจ อย่างมาก	นำพอ ใจ	ปาน กลาง	ไม่น่า พอใจ	ไม่น่าพอใจ อย่างมาก	นำพอ ใจ	ไม่น่า พอใจ
นำพอใจอย่างยิ่ง	√					√	
นำพอใจ	√					√	
ไม่น่าพอใจ	√						√
ไม่น่าพอใจ อย่างยิ่ง	√						√

จากตารางที่ 50 ผลการตัดสินคุณภาพการเปรียบเทียบคะแนนเมื่อใช้เกณฑ์ที่พัฒนาขึ้นเปรียบเทียบกับผลการใช้ตามเกณฑ์ของปีเตอร์เซนและคณะ พบว่าไม่สอดคล้องกัน โดยผลการตัดสินคุณภาพการเปรียบเทียบทั้ง 4 ระดับจากเกณฑ์ที่พัฒนาขึ้น อยู่ในระดับนำพอใจอย่างมาก เมื่อเทียบกับเกณฑ์ของปีเตอร์เซนและคณะ ส่วนผลการตัดสินคุณภาพการเปรียบเทียบคะแนนเมื่อใช้เกณฑ์ที่พัฒนาขึ้นเปรียบเทียบกับผลการใช้เกณฑ์ความเสมอภาคของลอร์ด พบว่ามีความสอดคล้องกัน โดยผลการตัดสินคุณภาพการเปรียบเทียบระดับนำพอใจอย่างยิ่งหรือระดับนำพอใจ และระดับไม่น่าพอใจหรือระดับไม่น่าพอใจอย่างยิ่งจากการใช้เกณฑ์ที่พัฒนาขึ้น อยู่ในระดับคุณภาพนำพอใจและระดับคุณภาพไม่น่าพอใจตามลำดับ เมื่อใช้เกณฑ์ความเสมอภาคของลอร์ด

ตารางที่ 51 ความสอดคล้องของเกณฑ์ที่พัฒนาขึ้นสำหรับเกณฑ์โมเดล 3 พารามิเตอร์ กับ  
เกณฑ์ความเสมอภาคของลอร์ด และเกณฑ์ของปีเตอร์เซนและคณะ

เกณฑ์ที่พัฒนาขึ้น	ระดับคุณภาพการเปรียบเทียบ						
	เกณฑ์ของปีเตอร์เซนและคณะ					เกณฑ์ของลอร์ด	
	นำพอใจ อย่างมาก	นำพอ ใจ	ปาน กลาง	ไม่น่า พอใจ	ไม่น่าพอใจ อย่างมาก	นำพอ ใจ	ไม่น่า พอใจ
นำพอใจอย่างยิ่ง	√					√	
นำพอใจ	√					√	
ไม่น่าพอใจ	√						√
ไม่น่าพอใจ อย่างยิ่ง	√						√

จากตารางที่ 51 ผลการตัดสินคุณภาพการเปรียบเทียบคะแนนเมื่อใช้เกณฑ์ที่พัฒนาขึ้นเปรียบเทียบกับผลการใช้ตามเกณฑ์ของปีเตอร์เซนและคณะ พบว่าไม่สอดคล้องกัน โดยผลการตัดสินคุณภาพการเปรียบเทียบทั้ง 4 ระดับจากเกณฑ์ที่พัฒนาขึ้น อยู่ในระดับน่าพอใจอย่างมาก เมื่อเทียบกับเกณฑ์ของปีเตอร์เซนและคณะ ส่วนผลการตัดสินคุณภาพการเปรียบเทียบคะแนนเมื่อใช้เกณฑ์ที่พัฒนาขึ้นเปรียบเทียบกับผลการใช้เกณฑ์ความเสมอภาคของลอร์ด พบว่ามีความสอดคล้องกัน โดยผลการตัดสินคุณภาพการเปรียบเทียบระดับน่าพอใจอย่างยิ่งหรือระดับน่าพอใจ และระดับไม่น่าพอใจหรือระดับไม่น่าพอใจอย่างยิ่งจากการใช้เกณฑ์ที่พัฒนาขึ้น อยู่ในระดับคุณภาพน่าพอใจและระดับคุณภาพไม่น่าพอใจตามลำดับ เมื่อใช้เกณฑ์ความเสมอภาคของลอร์ด

ตารางที่ 52 ความสอดคล้องของเกณฑ์ที่พัฒนาขึ้นสำหรับเกณฑ์แบบแผนเก็บรวบรวมข้อมูลใช้ข้อสอบร่วมกับเกณฑ์ความเสมอภาคของลอร์ด และเกณฑ์ของปีเตอร์เซนและคณะ

เกณฑ์ที่พัฒนาขึ้น	ระดับคุณภาพการเปรียบเทียบ						
	เกณฑ์ของปีเตอร์เซนและคณะ					เกณฑ์ของลอร์ด	
	น่าพอใจอย่างมาก	น่าพอใจ	ปานกลาง	ไม่น่าพอใจ	ไม่น่าพอใจอย่างมาก	น่าพอใจ	ไม่น่าพอใจ
น่าพอใจอย่างยิ่ง	√					√	
น่าพอใจ	√					√	
ไม่น่าพอใจ	√						√
ไม่น่าพอใจอย่างยิ่ง	√						√

จากตารางที่ 52 ผลการตัดสินคุณภาพการเปรียบเทียบคะแนนเมื่อใช้เกณฑ์ที่พัฒนาขึ้นเปรียบเทียบกับผลการใช้ตามเกณฑ์ของปีเตอร์เซนและคณะ พบว่าไม่สอดคล้องกัน โดยผลการตัดสินคุณภาพการเปรียบเทียบทั้ง 4 ระดับจากเกณฑ์ที่พัฒนาขึ้น อยู่ในระดับน่าพอใจอย่างมาก เมื่อเทียบกับเกณฑ์ของปีเตอร์เซนและคณะ ส่วนผลการตัดสินคุณภาพการเปรียบเทียบคะแนนเมื่อใช้เกณฑ์ที่พัฒนาขึ้นเปรียบเทียบกับผลการใช้เกณฑ์ความเสมอภาคของลอร์ด พบว่ามีความสอดคล้องกัน โดยผลการตัดสินคุณภาพการเปรียบเทียบระดับน่าพอใจอย่างยิ่งหรือระดับน่าพอใจ และระดับไม่น่าพอใจหรือระดับไม่น่าพอใจอย่างยิ่งจากการใช้เกณฑ์ที่พัฒนาขึ้น อยู่ในระดับคุณภาพน่าพอใจและระดับคุณภาพไม่น่าพอใจตามลำดับ เมื่อใช้เกณฑ์ความเสมอภาคของลอร์ด



ตารางที่ 53 ความสอดคล้องของเกณฑ์ที่พัฒนาขึ้นสำหรับเกณฑ์แบบแผนเก็บรวบรวมข้อมูลใช้  
กลุ่มสมมูล กับเกณฑ์ความเสมอภาคของลอร์ด และเกณฑ์ของปีเตอร์เซนและคณะ

เกณฑ์ที่พัฒนาขึ้น	ระดับคุณภาพการเปรียบเทียบ						
	เกณฑ์ของปีเตอร์เซนและคณะ					เกณฑ์ของลอร์ด	
	นำพอใจ อย่างมาก	นำพอ ใจ	ปาน กลาง	ไม่น่า พอใจ	ไม่น่าพอใจ อย่างมาก	นำพอ ใจ	ไม่น่า พอใจ
นำพอใจอย่างยิ่ง	√					√	
นำพอใจ	√					√	
ไม่น่าพอใจ	√						√
ไม่น่าพอใจ อย่างยิ่ง	√						√

จากตารางที่ 53 ผลการตัดสินคุณภาพการเปรียบเทียบคะแนนเมื่อใช้เกณฑ์ที่พัฒนา  
ขึ้นเปรียบเทียบกับผลการใช้ตามเกณฑ์ของปีเตอร์เซนและคณะ พบว่าไม่สอดคล้องกัน โดยผล  
การตัดสินคุณภาพการเปรียบเทียบทั้ง 4 ระดับจากเกณฑ์ที่พัฒนาขึ้น อยู่ในระดับนำพอใจอย่างมาก  
เมื่อเทียบกับเกณฑ์ของปีเตอร์เซนและคณะ ส่วนผลการตัดสินคุณภาพการเปรียบเทียบคะแนนเมื่อ  
ใช้เกณฑ์ที่พัฒนาขึ้นเปรียบเทียบกับผลการใช้เกณฑ์ความเสมอภาคของลอร์ด พบว่ามีความ  
สอดคล้องกัน โดยผลการตัดสินคุณภาพการเปรียบเทียบระดับนำพอใจอย่างยิ่งหรือระดับนำพอใจ  
และระดับไม่น่าพอใจหรือระดับไม่น่าพอใจอย่างยิ่งจากการใช้เกณฑ์ที่พัฒนาขึ้น อยู่ในระดับ  
คุณภาพนำพอใจและระดับคุณภาพไม่น่าพอใจตามลำดับ เมื่อใช้เกณฑ์ความเสมอภาคของลอร์ด

ตารางที่ 54 ความสอดคล้องของเกณฑ์ที่พัฒนาขึ้นสำหรับเกณฑ์ที่หาคุณภาพการปรับเทียบด้วย  
การปรับเทียบกลับสู่แบบสอบเดิม กับเกณฑ์ความเสมอภาคของลอร์ด และเกณฑ์  
ของปีเตอร์เซนและคณะ

เกณฑ์ที่พัฒนาขึ้น	ระดับคุณภาพการปรับเทียบ						
	เกณฑ์ของปีเตอร์เซนและคณะ					เกณฑ์ของลอร์ด	
	นำพอใจ อย่างมาก	นำพอ ใจ	ปาน กลาง	ไม่น่า พอใจ	ไม่น่าพอใจ อย่างมาก	นำพอ ใจ	ไม่น่า พอใจ
นำพอใจอย่างยิ่ง	√					√	
นำพอใจ	√					√	
ไม่น่าพอใจ	√						√
ไม่น่าพอใจ อย่างยิ่ง	√						√

จากตารางที่ 54 ผลการตัดสินคุณภาพการเปรียบเทียบคะแนนเมื่อใช้เกณฑ์ที่พัฒนาขึ้นเปรียบเทียบกับผลการใช้ตามเกณฑ์ของปีเตอร์เซนและคณะ พบว่าไม่สอดคล้องกัน โดยผลการตัดสินคุณภาพการเปรียบเทียบทั้ง 4 ระดับจากเกณฑ์ที่พัฒนาขึ้น อยู่ในระดับน่าพอใจอย่างมากเมื่อเทียบกับเกณฑ์ของปีเตอร์เซนและคณะ ส่วนผลการตัดสินคุณภาพการเปรียบเทียบคะแนนเมื่อใช้เกณฑ์ที่พัฒนาขึ้นเปรียบเทียบกับผลการใช้เกณฑ์ความเสมอภาคของลอร์ด พบว่ามีความสอดคล้องกัน โดยผลการตัดสินคุณภาพการเปรียบเทียบระดับน่าพอใจอย่างยิ่งหรือระดับน่าพอใจ และระดับไม่น่าพอใจหรือระดับไม่น่าพอใจอย่างยิ่งจากการใช้เกณฑ์ที่พัฒนาขึ้น อยู่ในระดับคุณภาพน่าพอใจและระดับคุณภาพไม่น่าพอใจตามลำดับ เมื่อใช้เกณฑ์ความเสมอภาคของลอร์ด

ตารางที่ 55 ความสอดคล้องของเกณฑ์ที่พัฒนาขึ้นสำหรับเกณฑ์ที่หาคุณภาพการปรับเทียบด้วยการใช้กลุ่มสอบทานผล กับเกณฑ์ความเสมอภาคของลอร์ด และเกณฑ์ของปีเตอร์เซนและคณะ

เกณฑ์ที่พัฒนาขึ้น	ระดับคุณภาพการปรับเทียบ						
	เกณฑ์ของปีเตอร์เซนและคณะ					เกณฑ์ของลอร์ด	
	น่าพอใจอย่างมาก	น่าพอใจ	ปานกลาง	ไม่น่าพอใจ	ไม่น่าพอใจอย่างมาก	น่าพอใจ	ไม่น่าพอใจ
น่าพอใจอย่างยิ่ง	√					√	
น่าพอใจ	√					√	
ไม่น่าพอใจ	√						√
ไม่น่าพอใจอย่างยิ่ง	√						√

จากตารางที่ 55 ผลการตัดสินคุณภาพการปรับเทียบคะแนนเมื่อใช้เกณฑ์ที่พัฒนาขึ้นเปรียบเทียบกับผลการใช้ตามเกณฑ์ของปีเตอร์เซนและคณะ พบว่าไม่สอดคล้องกัน โดยผลการตัดสินคุณภาพการปรับเทียบทั้ง 4 ระดับจากเกณฑ์ที่พัฒนาขึ้น อยู่ในระดับน่าพอใจอย่างมากเมื่อเทียบกับเกณฑ์ของปีเตอร์เซนและคณะ ส่วนผลการตัดสินคุณภาพการปรับเทียบคะแนนเมื่อใช้เกณฑ์ที่พัฒนาขึ้นเปรียบเทียบกับผลการใช้เกณฑ์ความเสมอภาคของลอร์ด พบว่ามีความสอดคล้องกัน โดยผลการตัดสินคุณภาพการปรับเทียบระดับน่าพอใจอย่างยิ่งหรือระดับน่าพอใจ และระดับไม่น่าพอใจหรือระดับไม่น่าพอใจอย่างยิ่งจากการใช้เกณฑ์ที่พัฒนาขึ้น อยู่ในระดับคุณภาพน่าพอใจและระดับคุณภาพไม่น่าพอใจตามลำดับ เมื่อใช้เกณฑ์ความเสมอภาคของลอร์ด

เกณฑ์ที่พัฒนาขึ้นเป็นเกณฑ์ตัดสินคุณภาพการเปรียบเทียบตามทฤษฎีการตอบสนองข้อสอบ มี 7 เกณฑ์ คือ เกณฑ์รวมเงื่อนไขทั้งหมด เกณฑ์สำหรับโมเดล 1 พารามิเตอร์ เกณฑ์สำหรับโมเดล 3 พารามิเตอร์ เกณฑ์สำหรับแบบแผนการเก็บรวบรวมข้อมูลใช้ข้อสอบร่วม เกณฑ์สำหรับแบบแผนการเก็บรวบรวมข้อมูลใช้กลุ่มสมมูล เกณฑ์สำหรับการหาคุณภาพการเปรียบเทียบคะแนนด้วยการเปรียบเทียบกลับสู่แบบสอบเดิม เกณฑ์สำหรับการหาคุณภาพการเปรียบเทียบการใช้กลุ่มสอบทานผล แต่ละเกณฑ์เป็นช่วงดัชนีความแตกต่าง 4 ช่วงที่บอกระดับคุณภาพการเปรียบเทียบคะแนน ทุกเกณฑ์เป็นเกณฑ์ที่มีคุณภาพเพราะเป็นเกณฑ์ที่ได้จากเงื่อนไขที่เป็นไปได้เกือบทั้งหมดในการเปรียบเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ มีกระบวนการกำหนดจุดตัดด้วยสถิติที่มีความเชื่อถือได้ โดยใช้ค่าดัชนีความแตกต่างระหว่างคะแนนจริงที่ไม่ได้ปรับเทียบกับคะแนนจริงที่ปรับเทียบแล้วเมื่อทดสอบความแตกต่างของค่าเฉลี่ยด้วย t-test แบบ two Dependent Sample Test แล้วพบว่าค่า t คำนวณเท่ากับค่าวิกฤต เมื่อเทียบเกณฑ์ที่พัฒนาขึ้นกับเกณฑ์ที่นิยมใช้กันมากคือเกณฑ์ของปีเตอร์เซนและคณะ ปรากฏว่าผลการตัดสินคุณภาพการเปรียบเทียบคะแนนไม่สอดคล้องกัน และเมื่อผลการตัดสินคุณภาพทั้ง ทั้ง 4 ระดับ คือ ระดับนำพอใจอย่างยิ่ง ระดับนำพอใจ ระดับไม่นำพอใจ และระดับไม่นำพอใจอย่างยิ่ง จะอยู่ในระดับนำพอใจอย่างยิ่งเมื่อใช้เกณฑ์ของปีเตอร์เซนและคณะ และเมื่อเทียบเกณฑ์ที่พัฒนาขึ้นกับเกณฑ์ความเสมอภาคของลอร์ด พบว่ามีความสอดคล้องกัน โดยผลการตัดสินคุณภาพการเปรียบเทียบระดับนำพอใจอย่างยิ่งหรือระดับนำพอใจและระดับไม่นำพอใจหรือระดับไม่นำพอใจอย่างยิ่งจากการใช้เกณฑ์ที่พัฒนาขึ้น อยู่ในระดับคุณภาพนำพอใจและระดับคุณภาพไม่นำพอใจตามลำดับเมื่อใช้เกณฑ์ความเสมอภาคของลอร์ด