

การเปรียบเทียบประสิทธิภาพของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในแบบทดสอบ
ที่มีการให้คะแนนแบบหลายค่าระหว่างวิธี GMH และวิธี Polytomous SIBTEST

ปริญญานิพนธ์
ของ
อุทัยวรรณ สายพัฒนา

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษาดุษฎีบัณฑิต สาขาการทดสอบและวัดผลการศึกษา
มีนาคม 2547
ลิขสิทธิ์เป็นของมหาวิทยาลัยศรีนครินทรวิโรฒ

371.15

08.1.13

1-3

การเปรียบเทียบประสิทธิภาพของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในแบบทดสอบ
ที่มีการให้คะแนนแบบหลายค่าระหว่างวิธี GMH และวิธี Polytomous SIBTEST

บทคัดย่อ

ของ

อุทัยวรรณ สายพัฒนา

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษาคุณวุฒิบัณฑิต สาขาการทดสอบและวัดผลการศึกษา
มีนาคม 2547

อุทัยวรรณ สายพัฒนา (2547). การเปรียบเทียบประสิทธิภาพของผลการตรวจสอบการทำหน้าที่ต่างกันของ
ข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่า ระหว่างวิธี GMH และวิธี Polytomous SIBTEST.
ปริญญาณิพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัย
ศรีนครินทรวิโรฒ, คณะกรรมการควบคุม: รองศาสตราจารย์ ดร.บุญเชิด ภิญโญนนตพงษ์, อาจารย์
ดร.ชูศักดิ์ ชัมภลสิทธิ์, อาจารย์ ดร.ละเอียด รักษ์เฒ่า.

การศึกษาค้นคว้าครั้งนี้มีจุดมุ่งหมายเพื่อเปรียบเทียบประสิทธิภาพของวิธี GMH และวิธี Polytomous SIBTEST
ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ จากข้อมูลผลการตอบข้อสอบที่มีการให้คะแนนแบบหลายค่า
ในเงื่อนไขความยาวของแบบทดสอบขนาด 20 30 และ 40 ข้อ กลุ่มตัวอย่างขนาด 250 500 และ 1,000 คน
กลุ่มตัวอย่างในการศึกษาได้มาจากการสุ่มแบบไล่คืบจากประชากรเทียม ซึ่งกำหนดจากนักเรียนชายและนักเรียน
หญิง ชั้นมัธยมศึกษาปีที่ 3 ในสังกัดกรมสามัญศึกษา จังหวัดพระนครศรีอยุธยา แต่ละขนาดกลุ่มตัวอย่างสุ่ม
จำนวน 10 กลุ่ม เครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบวิชาคณิตศาสตร์ เรื่อง วงกลม ชั้นมัธยมศึกษาปีที่ 3
ที่ผู้วิจัยสร้างขึ้น ประกอบด้วยข้อสอบเลือกตอบแบบถูกผิด ชนิด 5 ตัวเลือก จำนวน 40 ข้อ มีข้อสอบที่ผู้เชี่ยวชาญ
พิจารณาว่าเป็นข้อสอบที่ทำหน้าที่ต่างกันเมื่อต่อเพศชาย จำนวน 8 ข้อ หลังจากเก็บรวบรวมข้อมูลแล้ว คัดเลือก
ข้อสอบตามสัดส่วนในตารางกำหนดข้อสอบ จัดเป็นแบบทดสอบที่มีความยาว 30 ข้อและ 20 ข้อ แล้วตรวจสอบ
การทำหน้าที่ต่างกันของข้อสอบในเงื่อนไขต่าง ๆ ด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูป EXCEL และ MATLAB
สำหรับวิธี GMH และโปรแกรมสำเร็จรูป DIFPACK โปรแกรมย่อย Poly-SIBTEST สำหรับวิธี Polytomous
SIBTEST จากนั้น นำผลการวิเคราะห์ที่ได้มาหาความถูกต้องและความผิดพลาดในการตรวจสอบการทำหน้าที่
ต่างกันของข้อสอบ ด้วยการเปรียบเทียบกับผลที่ได้จากการตรวจสอบด้วยวิธี GPCM แล้วนำผลที่ได้ไปเปรียบ
เทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างขนาดของกลุ่มตัวอย่างของแต่ละวิธี
และต่างวิธีกัน โดยการใช้การวิเคราะห์ความแปรปรวนแบบหลายตัวแปร (MANOVA)

ผลการวิจัยสรุปได้ดังนี้

1. การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างขนาดของ
กลุ่มตัวอย่างของแต่ละวิธีการ ได้ผลดังนี้

1.1 วิธี GMH เมื่อแบบทดสอบประกอบด้วยข้อสอบ 40 30 และ 20 ข้อ กลุ่มตัวอย่างขนาด
1,000, 500 และ 250 คน ส่งผลต่อความถูกต้องและความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของ
ข้อสอบแตกต่างกัน ยกเว้น กรณีแบบทดสอบประกอบด้วยข้อสอบ 30 ข้อและ 20 ข้อ กลุ่มตัวอย่างขนาด 500
และ 250 คน ส่งผลต่อความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ไม่แตกต่างกัน

1.2 วิธี Polytomous SIBTEST เมื่อแบบทดสอบประกอบด้วยข้อสอบ 40, 30 และ 20 ข้อ กลุ่ม
ตัวอย่างขนาด 1,000, 500 และ 250 คน ส่งผลต่อความถูกต้องและความผิดพลาดในการตรวจสอบการทำ
หน้าที่ต่างกันของข้อสอบด้วยแตกต่างกัน ยกเว้น กรณีแบบทดสอบประกอบด้วยข้อสอบ 20 ข้อ กลุ่มตัวอย่าง
ขนาด 1,000 และ 500 คนส่งผลต่อความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบไม่แตกต่างกัน

2. การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี GMH และ
วิธี Polytomous SIBTEST ในทุกเงื่อนไขความยาวของแบบทดสอบ กลุ่มตัวอย่างทุกขนาดส่งผลต่อประสิทธิภาพ
ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของวิธีการทั้งสองไม่แตกต่างกัน ยกเว้นกรณีกลุ่มตัวอย่างขนาด
1,000 คน ของแบบทดสอบขนาด 20 ข้อ วิธี Polytomous SIBTEST มีประสิทธิภาพในการตรวจสอบการทำ
หน้าที่ต่างกันของข้อสอบสูงกว่าวิธี GMH

A COMPARISON OF THE EFFICIENCY OF GMH AND POLYTOMOUS SIBTEST
PROCEDURES ON DETECTING DIFFERENTIAL ITEM FUNCTIONING
IN POLYTOMOUSLY SCORED ITEMS

AN ABSTRACT
BY
UTHAIWAN SAIPATTHANA

Presented in partial fulfillment of the requirements
for the Doctor of Education degree in Testing and Measurement
at Srinakharinwirot University
March 2004

Uthaiwan Saipattana. (2004). A Comparison of the Efficiency of GMH and Polytomous SIBTEST Procedures on Detecting Differential Item Functioning in Polytomously Scored Items. Dissertation, Ed.D. (Testing and Measurement). Bangkok: Graduate School, Srinakharinwirot University. Advisor Committee : Assoc. Prof. Dr.Boonchird Pinyoanuntapong, Dr.Choosak Khampalikit, Dr.La-iad Ruckpau.

The purpose of this study was to compare the efficiency of GMH procedure with that of Polytomous SIBTEST on detecting differential item functioning (DIF) based on polytomously scored item response data. The three test lengths were of 20, 30 and 40 items and the three sample sizes were of 250, 500 and 1,000 students. The samples were selected by random sampling from pseudo population, consisting of Mathayomsuksa III, male and female students of Pranakomsriyutthaya province's secondary schools, under the jurisdiction of the Department of General Education. Ten groups for each sample size were taken as the sampling. The research instrument was a Mathayomsuksa III mathematics test on the circle area, constructed by the researcher. The test was five choices of Multiple True-False Items comprising 40 items, 8 of which were identified on experts' opinions as differential functioning in favor of male. After the data were collected, the test items were selected proportionately, based on the table of specification and categorized into the 30-item and 20-item tests. After that, the differential item functioning was all analyzed by EXCEL and MATLAB computer programs for GMH procedure and by Poly-SIBTEST subprogram in DIFPACK computer program for Polytomous SIBTEST procedure. Then the results were taken to find the consistency with the result of GPCM procedure, for the power of test and the type I error of the DIF detection. Lastly, the multivariate analysis of variance was used to compare the efficiency of the DIF detection results within and between the two procedures.

Results of the study are summarized as follows:

1. The comparison of the efficiency of the DIF detection results of each procedure among different sample sizes revealed the following :

- 1.1 When GMH procedure was used for combinations of the test lengths of 40, 30 and 20 items and the sample sizes of 1,000, 500 and 250 students, the power of test and the type I error of DIF detection were significantly different. However, at the 30 and 20 item tests for 500 and 250 students, the type I error of DIF detection was not different.

- 1.2 When Polytomous SIBTEST procedure was used for combinations of the test lengths of 40, 30 and 20 items and the sample sizes of 1,000, 500 and 250 students, the power of test and the type I error of DIF detection were significantly different. However, at the 20 item test for 1,000 and 500 students, the type I error of DIF detection was not different.

2. For all combinations of the test lengths and sample sizes, the efficiency of DIF detections between GMH and Polytomous SIBTEST procedures were not different. However, at the sample size of 1,000 students and the 20 item test, the efficiency of DIF detection for Polytomous SIBTEST procedure was more significantly than GMH procedure.

ปริญญานิพนธ์
เรื่อง

การเปรียบเทียบประสิทธิภาพของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ
ในแบบทดสอบที่มีการให้คะแนนแบบหลายค่าระหว่างวิธี GMH และวิธี Polytomous SIBTEST

ของ
นางสาวอุทัยวรรณ สายพัฒนา

ได้รับอนุมัติจากบัณฑิตวิทยาลัยให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาการศึกษาดุริยบัณฑิต สาขาการทดสอบและวัดผลการศึกษา
ของมหาวิทยาลัยศรีนครินทรวิโรฒ

.....คณบดีบัณฑิตวิทยาลัย
(รองศาสตราจารย์ ดร.นภาพร หะวานนท์)
วันที่ 10 เดือน มีนาคม พ.ศ. 2547

คณะกรรมการสอบปริญญานิพนธ์

.....ประธาน
(รองศาสตราจารย์ ดร.บุญเชิด ภิญโญนันตพงษ์)

.....กรรมการ
(อาจารย์ ดร.ชูศักดิ์ ชัมภลขิต)

.....กรรมการ
(อาจารย์ ดร.ละเอียต รักษ์เฝ้า)

.....กรรมการที่แต่งตั้งเพิ่มเติม
(รองศาสตราจารย์ ดร.สมสรร วงษ์อยู่น้อย)

.....กรรมการที่แต่งตั้งเพิ่มเติม
(อาจารย์ ดร.รังสรรค์ มณีเล็ก)

ประกาศคุณูปการ

ปริญญานิพนธ์ฉบับนี้สำเร็จสมบูรณ์ได้ด้วยความกรุณาและการช่วยเหลือจากคณะกรรมการที่ปรึกษาปริญญานิพนธ์ ทั้ง 3 ท่าน คือ รองศาสตราจารย์ ดร.บุญเชิด ภิญโญนนท์พงษ์ อาจารย์ ดร.ชูศักดิ์ ชัมภลขิต และอาจารย์ ดร.ละเอียด รัชเณร ที่ให้คำปรึกษา แนะนำแก้ไขข้อบกพร่องต่าง ๆ และให้กำลังใจแก่ผู้วิจัยด้วยดี ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูงมา ณ โอกาสนี้ และขอขอบพระคุณคณะกรรมการสอบเค้าโครงปริญญานิพนธ์และคณะกรรมการสอบปริญญานิพนธ์ทุกท่านที่กรุณาให้คำแนะนำที่เป็นประโยชน์ ทำให้ปริญญานิพนธ์ฉบับนี้มีความสมบูรณ์และมีคุณค่ายิ่งขึ้น และขอขอบคุณบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ที่ให้ทุนสนับสนุนส่วนหนึ่งในการทำวิจัยครั้งนี้

ขอขอบคุณ Prof. Dr. William Stout แห่งมหาวิทยาลัยจอร์เจียและ Dr. Xueli Xu แห่งมหาวิทยาลัยอิลลินอยส์ ที่ให้คำแนะนำเกี่ยวกับโปรแกรมคอมพิวเตอร์ในการวิเคราะห์ข้อมูล (DIFPACK) ขอขอบคุณอาจารย์ ดร.สุพจน์ เกิดสุวรรณ ที่กรุณาอนุเคราะห์โปรแกรมคอมพิวเตอร์สำเร็จรูปในการวิเคราะห์ข้อมูล (PARSCALE) พร้อมทั้งให้คำแนะนำเกี่ยวกับการใช้โปรแกรม ขอขอบคุณอาจารย์ปรีชา วิจิตรธรรมรส แห่งสถาบันบัณฑิตพัฒนบริหารศาสตร์ (NIDA) ที่กรุณาให้คำปรึกษา คำแนะนำ เกี่ยวกับวิธีการวิเคราะห์ GMH ทำให้ผู้วิจัยได้เรียนรู้แนวทางการวิเคราะห์ ขอขอบคุณอาจารย์ศิริณา บุญยสงวน แห่งคณะมนุษยศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ที่กรุณาตรวจสอบความถูกต้องของบทคัดย่อ (ภาคภาษาอังกฤษ) และขอขอบคุณอาจารย์ ดร.องอาจ นัยพัฒน์ ที่ได้อนุเคราะห์โปรแกรมคอมพิวเตอร์สำเร็จ รูป MATLAB ทำให้ผู้วิจัยสามารถวิเคราะห์ข้อมูลได้สะดวกและง่ายขึ้น ขอขอบคุณคุณฉัตรศิริ ปิยะพิมลสิทธิ์ ที่ช่วยเหลือ แก้ปัญหาเกี่ยวกับคอมพิวเตอร์

ขอขอบคุณคณะผู้เชี่ยวชาญ อาจารย์พรณี อุ่นศรีเพ็ญ อาจารย์กรรณิกา สุภาภา อาจารย์สุพรรณ บุรณพงศ์ อาจารย์วรพร สกฤตจิตรานนท์ และอาจารย์เบญจา เขียวสม ที่กรุณาตรวจสอบเครื่องมือวิจัย ให้คำแนะนำ และข้อคิดเห็นต่าง ๆ ในการแก้ไขและปรับปรุงเครื่องมือที่ใช้ในการวิจัยครั้งนี้

ขอขอบคุณผู้บริหาร คณะครู อาจารย์ทุกท่านของโรงเรียนที่เป็นกลุ่มตัวอย่าง และขอขอบคุณนักเรียนทุกคนที่ให้ความร่วมมือในการเก็บรวบรวมข้อมูลเป็นอย่างดี

ขอขอบคุณ คุณเบญจา เขียวสม คุณสุนิสา จั๊ยม่วงศรี และ คุณสุวรรณณี ละออบักษิน รวมทั้ง กัลยาณมิตรทุกคนที่คอยช่วยเหลือและให้กำลังใจแก่ผู้วิจัยด้วยดีเสมอมา

กราบขอบพระคุณ บิดา มารดา ผู้คอยห่วงใย เป็นกำลังใจ ให้การช่วยเหลือ และปรารถนาจะเห็นความสำเร็จของผู้วิจัยเป็นอย่างยิ่ง และกราบขอบพระคุณครู อาจารย์ทุกท่านทั้งในอดีตและปัจจุบันที่ได้อบรมสั่งสอนและประสิทธิ์ประสาทวิชาความรู้ จนผู้วิจัยประสบความสำเร็จในการศึกษาและหน้าที่การงานด้วยดี

คุณค่าและประโยชน์อันพึงมีจากปริญญานิพนธ์ฉบับนี้ ผู้วิจัยขอมอบเป็นเครื่องบูชาพระคุณ บิดา มารดา และครู อาจารย์ ตลอดจนผู้มีพระคุณต่อผู้วิจัยทุกท่าน

อุทัยวรรณ สายพัฒนา

สารบัญ

บทที่	หน้า
1 บทนำ.....	1
ภูมิหลัง	1
ความมุ่งหมายของการวิจัย	7
ความสำคัญของการวิจัย	7
ขอบเขตของการวิจัย	7
นิยามศัพท์เฉพาะ	8
กรอบความคิดในการวิจัย	10
2 เอกสารงานวิจัยที่เกี่ยวข้อง.....	11
แนวคิดเกี่ยวกับการตรวจสอบการทำหน้าที่ต่างกัน	11
การตรวจสอบการทำหน้าที่ต่างกันด้วยวิธี GMH	13
การตรวจสอบการทำหน้าที่ต่างกันด้วยวิธี Polytomous SIBTEST	19
การตรวจสอบการทำหน้าที่ต่างกันด้วยวิธี GPCM	26
เอกสารและงานวิจัยที่เกี่ยวข้อง	40
3 วิธีดำเนินการวิจัย.....	46
แหล่งข้อมูล.....	46
ประชากร	46
กลุ่มตัวอย่าง	46
เครื่องมือที่ใช้ในการวิจัย	47
การสร้างเครื่องมือในการวิจัย	48
ลักษณะของเครื่องมือที่ใช้ในการวิจัย.....	53
การเก็บรวบรวมข้อมูล	54
การเตรียมข้อมูล	54
การวิเคราะห์ข้อมูล.....	56
4 ผลการวิเคราะห์ข้อมูล.....	61
ตอนที่ 1 การวิเคราะห์ข้อมูลเบื้องต้น.....	63
1.1 ค่าสถิติพื้นฐานของข้อมูลผลการสอบ และคุณภาพของแบบทดสอบ	63
1.2 การตรวจสอบความเป็นมิตติเดียวของแบบทดสอบ	64
ตอนที่ 2 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GPCM	66
ตอนที่ 3 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GMH และ วิธี Polytomous SIBTEST	84
ตอนที่ 3 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GMH และ วิธี Polytomous SIBTEST	84

สารบัญ (ต่อ)

บทที่	หน้า
5 สรุป อภิปรายผล และข้อเสนอแนะ.....	97
วิธีดำเนินการวิจัย.....	98
สรุปผลการวิจัย.....	99
อภิปรายผลการวิจัย.....	104
ข้อเสนอแนะ.....	112
บรรณานุกรม	114
ภาคผนวก	123
ประวัติย่อผู้วิจัย	333

บัญชีตาราง

ตาราง	หน้า
1 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่า จำแนกตาม ทฤษฎีที่ใช้ในการวิเคราะห์และรูปแบบของโมเดล.....	13
2 แสดงผลการตอบข้อสอบของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจที่ระดับคะแนน j	15
3 แสดงสัดส่วนของผลการตอบของผู้ตอบ ที่ระดับคะแนน j	15
4 แสดงผลการตอบของกลุ่มผู้สอบ ในแต่ละลำดับขั้นคะแนน	18
5 เปรียบเทียบวิธีการในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบ หลายค่า.....	39
6 แสดงรายชื่อโรงเรียน จำนวนห้องเรียน และจำนวนนักเรียน	47
7 ตารางกำหนดข้อสอบวิชาคณิตศาสตร์ เรื่องวงกลม จำแนกตามระดับพฤติกรรมที่วัด.....	51
8 ค่าสถิติพื้นฐานของแบบทดสอบวิชาคณิตศาสตร์ ขนาด 40 ข้อ, 30 ข้อและ 20 ข้อ.....	63
9 ค่าไอเกนและร้อยละของความแปรปรวนขององค์ประกอบของแบบทดสอบขนาด 40 ข้อ.....	64
10 ค่าไอเกนและร้อยละของความแปรปรวนขององค์ประกอบของแบบทดสอบขนาด 30 ข้อ.....	65
11 ค่าไอเกนและร้อยละของความแปรปรวนขององค์ประกอบของแบบทดสอบขนาด 20 ข้อ.....	66
12 ค่าพารามิเตอร์ของข้อสอบแยกตามกลุ่มเพศชาย-หญิง เมื่อกลุ่มตัวอย่าง 1,000 คน และ ความยาวของแบบทดสอบ 40 ข้อ	67
13 ค่าพารามิเตอร์ของข้อสอบแยกตามกลุ่มเพศชาย-หญิง เมื่อกลุ่มตัวอย่าง 1,000 คน และ ความยาวของแบบทดสอบ 30 ข้อ	68
14 ค่าพารามิเตอร์ของข้อสอบแยกตามกลุ่มเพศชาย-หญิง เมื่อกลุ่มตัวอย่าง 1,000 คน และ ความยาวของแบบทดสอบ 20 ข้อ	69
15 ค่าสัมประสิทธิ์การเชื่อมโยงของแบบทดสอบขนาด 40 ข้อ, 30 ข้อและ 20 ข้อ	70
16 ค่าพารามิเตอร์ของข้อสอบ ที่ปรับเทียบสเกลแล้ว แยกตามกลุ่มเพศชาย-หญิง เมื่อกลุ่มตัวอย่าง 1,000 คน และความยาวของแบบทดสอบ 40 ข้อ	71
17 ค่าพารามิเตอร์ของข้อสอบ ที่ปรับเทียบสเกลแล้ว แยกตามกลุ่มเพศชาย-หญิง เมื่อกลุ่มตัวอย่าง 1,000 คน และความยาวของแบบทดสอบ 30 ข้อ	72
18 ค่าพารามิเตอร์ของข้อสอบ ที่ปรับเทียบสเกลแล้ว แยกตามกลุ่มเพศชาย-หญิง เมื่อกลุ่มตัวอย่าง 1,000 คน และความยาวของแบบทดสอบ 20 ข้อ	73
19 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GPCM ตามแนวคิดของทฤษฎีการตอบ ข้อสอบด้วยวิธีของลอร์ด (Lord's χ^2) เมื่อกลุ่มตัวอย่าง 1,000 คนและความยาวของแบบ ทดสอบ 40 ข้อ.....	74
20 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GPCM ตามแนวคิดของทฤษฎีการตอบ ข้อสอบด้วยวิธีของลอร์ด (Lord's χ^2) เมื่อกลุ่มตัวอย่าง 1,000 คนและความยาวของแบบ ทดสอบ 30 ข้อ.....	76

บัญชีตาราง (ต่อ)

ตาราง	หน้า
21 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GPCM ตามแนวคิดของทฤษฎีการตอบข้อสอบด้วยวิธีของลอร์ด (Lord's χ^2) เมื่อกลุ่มตัวอย่าง 1,000 คนและความยาวของแบบทดสอบ 20 ข้อ.....	77
22 สรุปผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี GMH และวิธี Polytomous SIBTEST ในเงื่อนไขกลุ่มตัวอย่างและความยาวของแบบทดสอบขนาดแตกต่างกัน.....	84
23 ผลการทดสอบนัยสำคัญของขนาดกลุ่มตัวอย่างที่มีต่อความถูกต้องและความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ พร้อมกันและเป็นรายตัว ด้วยวิธี GMH	87
24 ผลการทดสอบความแตกต่างระหว่างค่าเฉลี่ยของความถูกต้องและความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GMH เมื่อแบบทดสอบและกลุ่มตัวอย่างขนาดต่างกัน	88
25 ผลการทดสอบนัยสำคัญของขนาดกลุ่มตัวอย่างที่มีต่อความถูกต้องและความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ พร้อมกันและเป็นรายตัวด้วยวิธี Polytomous SIBTEST	90
26 ผลการทดสอบความแตกต่างระหว่างค่าเฉลี่ยของความถูกต้องและความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี Polytomous SIBTEST เมื่อแบบทดสอบและกลุ่มตัวอย่างขนาดต่างกัน.....	91
27 ผลการทดสอบนัยสำคัญของวิธีการตรวจสอบที่มีต่อความถูกต้องและความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ พร้อมกัน.....	94
28 ผลการทดสอบความแตกต่างระหว่างค่าเฉลี่ยของความถูกต้องและความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ตามวิธี GMH และวิธี Polytomous SIBTEST ในเงื่อนไขขนาดกลุ่มตัวอย่างและความยาวของแบบทดสอบ ที่ต่างกัน.....	95
29 ผลการประเมินความสอดคล้องระหว่างเนื้อหากับระดับพฤติกรรมที่ต้องการ และการพิจารณาการทำหน้าที่ต่างกันของข้อสอบ.....	137
30 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) เมื่อกลุ่มตัวอย่างมีขนาด 1,000 คน และแบบทดสอบมีขนาด 20 ข้อ.....	139
31 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) เมื่อกลุ่มตัวอย่างมีขนาด 500 คน และแบบทดสอบมีขนาด 20 ข้อ.....	140
32 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) เมื่อกลุ่มตัวอย่างมีขนาด 250 คน และแบบทดสอบมีขนาด 20 ข้อ.....	141
33 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) เมื่อกลุ่มตัวอย่างมีขนาด 1,000 คน และแบบทดสอบมีขนาด 30 ข้อ.....	142
34 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) เมื่อกลุ่มตัวอย่างมีขนาด 500 คน และแบบทดสอบมีขนาด 30 ข้อ.....	143

บัญชีตาราง (ต่อ)

ตาราง	หน้า
35 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) เมื่อกลุ่มตัวอย่างมีขนาด 250 คน และ แบบทดสอบมีขนาด 30 ข้อ.....	144
36 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) เมื่อกลุ่มตัวอย่างมีขนาด 1,000 คน และ แบบทดสอบมีขนาด 40 ข้อ.....	145
37 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) เมื่อกลุ่มตัวอย่างมีขนาด 500 คน และ แบบทดสอบมีขนาด 40 ข้อ.....	146
38 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) เมื่อกลุ่มตัวอย่างมีขนาด 250 คน และ แบบทดสอบมีขนาด 40 ข้อ.....	147

บัญชีภาพประกอบ

ภาพประกอบ	หน้า
1 กรอบความคิดในการวิจัย.....	10
2 ขั้นตอนการสร้างเครื่องมือที่ใช้ในการวิจัย.....	48
3 กระบวนการเตรียมข้อมูลเพื่อใช้ในการวิเคราะห์.....	56
4 ผลการตรวจสอบความเป็นมิติเดียวของแบบทดสอบขนาด 40 ข้อ.....	64
5 ผลการตรวจสอบความเป็นมิติเดียวของแบบทดสอบขนาด 30 ข้อ.....	65
6 ผลการตรวจสอบความเป็นมิติเดียวของแบบทดสอบขนาด 20 ข้อ.....	66
7 โค้งคุณลักษณะข้อสอบ (ICC) ของข้อสอบที่มี DIF ของแบบทดสอบขนาด 20 ข้อ	79
8 โค้งคุณลักษณะข้อสอบ (ICC) ของข้อสอบที่มี DIF ของแบบทดสอบขนาด 30 ข้อ	80
9 โค้งคุณลักษณะข้อสอบ (ICC) ของข้อสอบที่มี DIF ของแบบทดสอบขนาด 40 ข้อ	81
10 ฟังก์ชันผลการตอบลำดับขั้นข้อสอบ (ICRF) ของข้อสอบข้อ 7 ของแบบทดสอบขนาด 20 ข้อ.....	82
11 ฟังก์ชันผลการตอบข้อสอบในขอบเขตลำดับขั้น(BRF) ของข้อสอบข้อ 7 ของแบบทดสอบขนาด 20 ข้อ.....	83
12 เกณฑ์การพิจารณาประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ.....	86

บทที่ 1

บทนำ

ภูมิหลัง

แบบทดสอบเลือกตอบ(Multiple-choice Test) เป็นเครื่องมือวัดผลที่นิยมใช้อย่างแพร่หลายมาก เนื่องจากสามารถวัดได้ครอบคลุมเนื้อหา มีความเป็นปรนัยสูงในการตรวจให้คะแนน ข้อคำถามและการแปลผลคะแนน สามารถนำมาวิเคราะห์หาคุณภาพ โดยเฉพาะมีความสะดวกในการดำเนินการสอบและการตรวจให้คะแนน แต่แบบทดสอบเลือกตอบที่ให้นักเรียนเลือกตอบเฉพาะตัวเลือกถูกเพียงตัวเลือกเดียวและมีการให้คะแนนแบบสองค่า (Dichotomously scored items) คือ ตอบถูกให้ 1 คะแนน ตอบผิดให้ 0 คะแนนนั้น นักวัดผลกลุ่มหนึ่งเกิดข้อสงสัยว่านักเรียนที่ตอบถูกนั้นเป็นผู้ที่มีความรู้ในเรื่องนั้นจริงหรือเกิดจากการเดาและนักเรียนที่ตอบผิดนั้นเป็นเพราะไม่มีความรู้ในเรื่องนั้นเลยหรือเกิดจากปัญหาอื่น ๆ จึงได้พยายามพัฒนาแบบทดสอบที่สามารถบ่งบอกความรู้แท้จริงของผู้เรียนให้มากที่สุดเท่าที่จะทำได้ กระบวนการพัฒนาแบบทดสอบเลือกตอบได้ดำเนินการอย่างต่อเนื่อง โดยเน้นที่วิธีการตอบและการให้คะแนน ซึ่งมีเป้าหมายเพื่อแก้จุดอ่อนของแบบทดสอบเลือกตอบเกี่ยวกับการเดาให้สามารถนำคะแนนผลการสอบไปประมาณค่าความเที่ยงตรงและความเชื่อมั่นของแบบทดสอบให้มีความคลาดเคลื่อนน้อยที่สุด แนวทางหนึ่งที่ใช้ คือ การใช้สูตรปรับคะแนนผลการสอบวัด ซึ่งเรียกว่า การปรับแก้คะแนนการเดา วิธีการดังกล่าวนี้จะเน้นที่คะแนนรวมทั้งฉบับ (Total Score) จึงไม่ให้รายละเอียดในส่วนที่เป็นรายข้อเท่าใดนัก ต่อมาจึงมีแนวทางใหม่เกิดขึ้นที่เหมาะสมกว่าด้วยการปรับวิธีการตอบและการให้คะแนนรายข้อใหม่ด้วยการปรับคะแนนแต่ละข้อ เป็นการให้คะแนนที่เน้นการแสดงความรู้อย่างบางส่วน (Partial knowledge) (Coombs, C.H. et al. 1956 : 13) และให้คะแนนแบบหลายค่า (Polytomous Score)

ข้อสอบชนิดให้คะแนนหลายค่า (Polytomously scored items) มีหลายประเภท เช่น ข้อสอบความเรียง (Essay items) การประเมินงานปฏิบัติ (Performance tasks assessment) อาทิจากการตัดสินคุณภาพของแฟ้มสะสมงาน (Portfolio) การพิสูจน์ทางคณิตศาสตร์ (Mathematical proof) การทดลองทางวิทยาศาสตร์ (Scientific experiment) เป็นต้น แบบทดสอบถูกผิดหลายตัวเลือก (Multiple True-False) เป็นแบบทดสอบประเภทหนึ่งที่สามารถให้คะแนนแบบหลายค่าได้ ซึ่งลักษณะของแบบทดสอบถูกผิดหลายตัวเลือกนั้น ผู้สอบต้องพิจารณาตัวเลือกทุกตัวเลือกว่าตัวเลือกใดถูกตัวเลือกใดผิด ซึ่งจำนวนตัวเลือกที่ถูกอาจมีตั้งแต่ 1 ตัวเลือกจนถึงถูกทุกตัวเลือก เพื่อเป็นการวัดว่าผู้ตอบมีความรู้ที่ถูกต้องทั้งหมดหรือไม่ ฉะนั้นแบบทดสอบถูกผิดหลายตัวเลือกจึงมีจำนวนตัวเลือกตั้งแต่สองตัวเลือกขึ้นไป ซึ่งควรมีมากกว่าสองตัวเลือก เพราะสามารถวัดว่านักเรียนมีความรู้ในเรื่องนั้นอย่างแท้จริงมากที่สุดและโอกาสในการเดาคำตอบจะมีน้อยลง (Downing, Steven M. et al. 1995 : 187 ; Albanese, Mark A. & Sabers, Darrell L. 1988 : 111)

การพัฒนาวิธีการตอบและการให้คะแนนแบบทดสอบเลือกตอบ จำเป็นต้องคำนึงถึงคุณภาพของแบบทดสอบ ซึ่งเป็นหัวใจของการวัดและการประเมินผล คุณภาพที่สำคัญที่สุดของแบบทดสอบ คือ ความเที่ยงตรง (Validity) เพราะการนำแบบทดสอบที่ขาดความเที่ยงตรงมาใช้ในการวัดเป็นการไม่ยุติธรรมสำหรับผู้สอบ ผลที่ได้จากการวัดไม่สามารถสะท้อนถึงความสามารถที่แท้จริงของผู้สอบ การตรวจสอบความเที่ยงตรงของแบบทดสอบสามารถทำได้หลายวิธี วิธีการหนึ่งคือการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ข้อสอบที่ไม่ได้วัดเฉพาะคุณลักษณะแฝง (Latent trait) ที่ต้องการวัด แต่วัดคุณลักษณะแฝงอื่นของผู้สอบที่ไม่ต้องการ

วัด แสดงว่า ข้อสอบนั้นขาดความเที่ยงตรง (Shealy & Stout.1993 : 197) เพราะถ้าผู้สอบกลุ่มย่อยใดมีคุณลักษณะแฝงอื่นสูงกว่าย่อมมีโอกาสที่จะตอบข้อสอบได้ถูกต้องมากกว่าทั้ง ๆ ที่มีคุณลักษณะแฝงที่ต้องการวัดเท่ากันกับกลุ่มผู้สอบย่อยกลุ่มอื่น จึงทำให้เกิดการได้เปรียบเสียเปรียบกันระหว่างกลุ่มผู้สอบย่อย ลักษณะนี้เดิมใช้คำว่า ความลำเอียงของข้อสอบ (Item Bias) ต่อมาระยะหลังเกิดความคลุมเครือในการที่จะใช้เกณฑ์ในการตัดสินใจเรื่องความลำเอียง จึงนิยมใช้สารสนเทศทางสถิติมาเป็นเกณฑ์ในการตัดสินใจ และใช้คำว่า การทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning : DIF) เนื่องจากเห็นว่าเป็นคำที่มีความหมายกลาง ๆ มีความเหมาะสมในเชิงวิชาการมากกว่าคำว่า ความลำเอียง (Holland & Wainer. 1993 : 4-5)

การทำหน้าที่ต่างกันของข้อสอบนั้น ขนาดและทิศทางของการทำหน้าที่ต่างกันของข้อสอบจะแปรเปลี่ยนไปตามระดับความสามารถที่แตกต่างกันของผู้สอบ โดยแบ่งลักษณะของข้อสอบที่ทำหน้าที่ต่างกันเป็น 2 ประเภท (Mellenbergh, 1982) ได้แก่ ข้อสอบที่ทำหน้าที่ต่างกันแบบสม่ำเสมอ (Uniform DIF) ซึ่งจะเกิดขึ้นเมื่อไม่มีปฏิสัมพันธ์ (interaction) ระหว่างระดับความสามารถของผู้สอบกับการเป็นสมาชิกของกลุ่มย่อย นั่นคือ โอกาสของการตอบข้อสอบได้ถูกต้องของผู้สอบกลุ่มย่อยกลุ่มหนึ่ง สูงกว่า ผู้สอบกลุ่มย่อยอีกกลุ่มหนึ่ง ตลอดช่วงความสามารถ และ ข้อสอบที่ทำหน้าที่ต่างกันแบบไม่สม่ำเสมอ (Nonuniform DIF) ซึ่งจะเกิดขึ้นเมื่อมีปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้สอบกับการเป็นสมาชิกของกลุ่มย่อย นั่นคือ โอกาสของการตอบข้อสอบได้ถูกต้องของผู้สอบกลุ่มย่อยกลุ่มหนึ่งสูงกว่าผู้สอบกลุ่มย่อยอีกกลุ่มหนึ่ง ไม่ตลอดช่วงความสามารถ ซึ่งตามทฤษฎีการตอบข้อสอบ (IRT) สามารถพิจารณาปฏิสัมพันธ์ได้จาก ความแตกต่างของค่าพารามิเตอร์อำนาจจำแนกของข้อสอบระหว่างผู้สอบกลุ่มย่อย 2 กลุ่ม ถ้าข้อสอบระหว่างผู้สอบ 2 กลุ่มย่อยมีค่าพารามิเตอร์อำนาจจำแนกเท่ากันแล้ว โค้งคุณลักษณะข้อสอบ (Item Characteristic Curves : ICC) ของผู้สอบ 2 กลุ่มจะขนานกัน แสดงว่า ข้อสอบที่ทำหน้าที่ต่างกันแบบสม่ำเสมอ แต่ถ้าข้อสอบระหว่างผู้สอบ 2 กลุ่มย่อยมีค่าพารามิเตอร์อำนาจจำแนกไม่เท่ากันแล้ว โค้งคุณลักษณะข้อสอบ (Item Characteristic Curves : ICC) ของผู้สอบ 2 กลุ่มจะไม่ขนานกัน แสดงว่า ข้อสอบที่ทำหน้าที่ต่างกันแบบไม่สม่ำเสมอ

วิธีการในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบสองค่า มีหลายวิธี จำแนกเป็น 2 กลุ่มวิธีใหญ่ ๆ คือ กลุ่มวิธี IRT และกลุ่มวิธีที่ไม่ใช่ IRT กลุ่มวิธี IRT เป็นกลุ่มที่วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบใช้คะแนนที่สังเกตไม่ได้ (unobserved score) หรือ ใช้ตัวแปรแฝง (latent variable) วิเคราะห์ตามทฤษฎีการตอบข้อสอบ (Item Response Theory : IRT) และใช้การประมาณค่าความสามารถของผู้สอบเป็นเกณฑ์การจับคู่กลุ่มผู้สอบ วิธีการตรวจสอบในกลุ่มนี้แบ่งเป็น 3 วิธีใหญ่ ๆ คือ วิธีการวัดพื้นที่ วิธีการเปรียบเทียบค่าพารามิเตอร์ และ วิธีซิปเทสต์ (SIBTEST) (Shealy and Stout. 1993) วิธีการวัดพื้นที่ ซึ่งแบ่งออกเป็นวิธีย่อย ๆ อีกหลายวิธี เช่น วิธีการวัดพื้นที่ของรูดเนอร์ (Rudner. 1977), วิธีการวัดพื้นที่ของลินน์ เลวิน ฮาส์ติงส์และวาร์ดริฟ (Linn, Levine, Hastings and Wardrup. 1981), วิธีการวัดพื้นที่ของเชพเพิร์ด คามิลลี และวิลเลียมส์ (Shepard, Camilli and Williams. 1984), วิธีการวัดพื้นที่ของราชู (Raju. 1990) และ วิธีการวัดพื้นที่ของคิมและโคเฮน (Kim and Cohen. 1991) เป็นต้น ส่วนวิธีการเปรียบเทียบค่าพารามิเตอร์ แบ่งออกเป็นวิธีย่อย ๆ เช่น วิธีการทดสอบไค - สแควร์ของลอร์ด (Lord's chi-square test) (1980) และวิธีการทดสอบอัตราส่วนความน่าจะเป็น (Likelihood Ratio test : LR) (Thissen, Steinberge and Wainer. 1993) เป็นต้น ส่วนกลุ่มวิธีที่ไม่ใช่ IRT เป็นกลุ่มที่วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบใช้คะแนนที่สังเกตได้ (observed score) วิเคราะห์ตามทฤษฎีการทดสอบมาตรฐานเดิม (Classical Test Theory : CTT) และ ใช้คะแนนรวมของผู้สอบเป็นเกณฑ์การจับคู่กลุ่มผู้สอบ ซึ่งเป็นเกณฑ์ภายใน วิธีการที่สำคัญในกลุ่มนี้ ได้แก่ วิธีแปลงค่าความยากของข้อสอบ (Transformed Item

Difficulty : TID) (Angoff, 1972 cited in Osterlind, 1983) วิธีการวิเคราะห์ความแปรปรวน (Analysis of Variance : ANOVA) (Cleary and Hillton, 1968 cited in Osterlind, 1983) วิธีการทำให้เป็นมาตรฐาน (standardization : STND) (Dorans and Kulick, 1986) วิธีล็อก-ลิเนียร์ (Log-Linear : LL) (Mellenbergh, 1982) วิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel : MH) (Holland and Thayer, 1988) และวิธีการถดถอยโลจิสติก (Logistic Regression : LR) (Swaminathan and Rogers, 1990) เป็นต้น

เนื่องจากการพัฒนาข้อสอบทั้งวิธีการตอบและการให้คะแนน ดังนั้น วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจึงมีการปรับเปลี่ยนวิธีการ และ พัฒนาให้สามารถใช้ได้กับข้อสอบที่มีการให้คะแนนแบบหลายค่าได้อย่างเหมาะสมและรัดกุมยิ่งขึ้น ซึ่ง โปเทนซ่าและดูรานส์ (Potenza & Durans, 1995) ได้รวบรวมวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ใช้กับข้อสอบที่มีการให้คะแนนแบบหลายค่า และ นำเสนอไว้ดังนี้ วิธีการตรวจสอบในกลุ่มวิธี IRT ได้แก่ วิธีการทดสอบอัตราส่วนความน่าจะเป็น (General IRT LR), วิธีพาเชียลเครดิตโมเดล (Partial Credit Model : PCM), วิธีชิปเทสต์แบบหลายค่า (Polytomous SIBTEST) และวิธีพาเชียลเครดิตโมเดลทั่วไป (Generalized Partial Credit Model : GPCM) เป็นต้น ส่วน วิธีการตรวจสอบใน กลุ่มวิธีที่ไม่ใช่ IRT ได้แก่ วิธีการทำให้เป็นมาตรฐานแบบหลายค่า (Polytomous STND), วิธีแมนเทล-แฮนส์เซลทั่วไป (Generalized Mantel-Haenszel : GMH) และวิธีการถดถอยโลจิสติกแบบหลายค่า (Polytomous LRDIF) นอกจากนั้น ยังมีวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก (Logistic Discriminant Function Analysis : LDFA) ซึ่งพัฒนาจากแนวคิดของวิธีการถดถอยโลจิสติกและวิธีการวิเคราะห์เชิงจำแนกกลุ่ม (Discriminant Analysis) เป็นต้น

วิธีพาเชียลเครดิตโมเดลทั่วไป (Generalized Partial Credit Model : GPCM) เป็นวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ข้อสอบมีการให้คะแนนแบบหลายค่า ภายใต้ทฤษฎีการตอบข้อสอบ (IRT) ซึ่งวิธีการตรวจสอบการทำหน้าที่ต่างกันที่ข้อสอบมีการให้คะแนนแบบสองค่า ภายใต้ทฤษฎี IRT เป็นที่ยอมรับกันโดยทั่วไปว่า มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูง เช่น เชปาร์ดและคณะ (Berk. 1982 : 169; citing Shepard, Camilli and Averill. 1980) ได้ทำการเปรียบเทียบวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 5 วิธี คือ วิธีไคสแควร์ วิธีแปลงค่าความยาก วิธีพอยท์ไบซีเรียล วิธีโด่งลักษณะข้อสอบ 3 พารามิเตอร์และวิธีโด่งลักษณะข้อสอบ 1 พารามิเตอร์ พบว่า วิธีโด่งลักษณะข้อสอบ 3 พารามิเตอร์ เป็นวิธีที่เหมาะสมในทางทฤษฎี มีความถูกต้องที่สุด ส่วนวิธีไคสแควร์ให้ผลสอดคล้องกับวิธีโด่งลักษณะข้อสอบ 3 พารามิเตอร์ สงบ ลักษณะ (Burrill. 1982 : 168 ; citing Laksana. 1979) ได้ศึกษาเปรียบเทียบวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 2 วิธี คือวิธีโด่งลักษณะข้อสอบ 3 พารามิเตอร์ และวิธีวิเคราะห์ความแปรปรวน พบว่า วิธีโด่งลักษณะข้อสอบ 3 พารามิเตอร์เป็นวิธีที่ให้ผลครอบคลุมถูกต้อง แต่วิธีวิเคราะห์ความแปรปรวนใช้ได้ดีในทางปฏิบัติ ส่วนรูดเนอร์และคนอื่น ๆ (Subkoviak. 1984 : 50; citing Rudner and others. 1978) ได้ศึกษาเปรียบเทียบวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 4 วิธี คือ วิธีแปลงค่าความยาก วิธีไคสแควร์ วิธีโด่งลักษณะข้อสอบ และ วิธีวิเคราะห์องค์ประกอบ พบว่า วิธีโด่งลักษณะข้อสอบและวิธีแปลงค่าความยากเป็นวิธีที่เหมาะสมที่สุด และจากงานวิจัยของ ทศนีย์ พิรมนตรี (2530 : 76-78) ได้ศึกษาเปรียบเทียบวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 3 วิธี คือวิธีกำหนดจุดเดลต้า วิธีไคสแควร์ และวิธีโด่งลักษณะข้อสอบ 3 พารามิเตอร์ พบว่า วิธีโด่งลักษณะข้อสอบ 3 พารามิเตอร์สามารถตรวจสอบข้อสอบทำหน้าที่ต่างกันมากที่สุด สุรศักดิ์ อมรรัตนศักดิ์ (2531 : 200-207) ได้ศึกษาเปรียบเทียบวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 4 วิธี คือ วิธีวิเคราะห์ความแปรปรวน วิธีแปลงค่าความยาก วิธีโด่งลักษณะข้อสอบ 1 พารามิเตอร์ และวิธีโด่งลักษณะข้อสอบ 3 พารามิเตอร์ พบว่า วิธีโด่ง

ลักษณะข้อสอบ 3 พารามิเตอร์ พบข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด จะเห็นได้ว่า ผลการวิจัยส่วนมากจะพบว่า วิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ ซึ่งเป็นวิธีการตามทฤษฎีการตอบข้อสอบนั้นให้ผลการตรวจสอบพบข้อสอบที่ทำหน้าที่ต่างกันถูกต้องมากที่สุด ทั้งนี้อาจเนื่องจากคุณลักษณะเด่นของทฤษฎีในหลาย ๆ ด้าน เช่น ค่าพารามิเตอร์ไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบ สามารถวิเคราะห์และอธิบายโอกาสในการตอบข้อสอบของผู้สอบที่มีความสามารถต่ำมาก ๆ ได้ ทำให้มีความไวในการตรวจสอบ สำหรับวิธี GPCM ซึ่งเป็นวิธีที่ใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนหลายค่า ยังสามารถพิจารณาการทำหน้าที่ต่างกันของข้อสอบในแต่ละลำดับชั้นการตอบ (Categories response) ได้ โดยพิจารณาจากฟังก์ชันผลการตอบลำดับชั้นข้อสอบ (Item Category Response Function : ICRF) และฟังก์ชันผลการตอบในขอบเขตลำดับชั้น (Boundary Response Function : BRF) อย่างไรก็ตาม แม้ว่าวิธีการตามทฤษฎีการตอบข้อสอบจะมีความเหมาะสมในทางทฤษฎีมากที่สุด แต่ปัญหาอุปสรรคของวิธีการนี้มีหลายประการที่ทำให้ไม่เหมาะสมในเชิงปฏิบัติ เช่น ความยุ่งยาก ซับซ้อนในการคิดคำนวณ ต้องใช้กลุ่มตัวอย่างขนาดใหญ่ ทำให้เสียค่าใช้จ่ายสูง ดังนั้น ควรศึกษาวิธีการอื่น ๆ ที่ให้ผลการตรวจสอบที่สอดคล้อง ถูกต้องใกล้เคียงกับวิธีการตามทฤษฎีการตอบข้อสอบ มีความสะดวกง่าย ประหยัดและมีความเหมาะสมในทางปฏิบัติ

วิธีแมนเทล-แฮนส์เซลเดิม ซึ่งพัฒนาโดย ฮอลแลนด์และทาเยอร์ (Holland & Thayer, 1988) เป็นวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในรูปแบบนันทพารามิเตอร์ (nonparametric) ไม่จำเป็นต้องใช้โมเดลประมาณค่าพารามิเตอร์ เป็นวิธีการที่มีเงื่อนไข (condition methods) คือ ใช้หลักการจับคู่ผู้สอบของกลุ่มอ้างอิงกับกลุ่มสนใจตามความสามารถของผู้สอบ (Somes, 1986: 107) มีการทดสอบนัยสำคัญและมีการวัดขนาดอิทธิพล เทคนิคทางสถิติมีความเชื่อถือได้ สามารถคำนวณได้ง่าย ไม่ยุ่งยากซับซ้อน มีหลักการที่สามารถเข้าใจได้ง่าย เสียค่าใช้จ่ายน้อย และใช้ได้กับกลุ่มตัวอย่างขนาดเล็ก เมื่อเทียบกับวิธีการที่ใช้ทฤษฎีการตอบข้อสอบและผลการตรวจสอบพบข้อสอบที่ทำหน้าที่ต่างกันใกล้เคียงกับวิธีที่ใช้ทฤษฎีการตอบข้อสอบ (Hambleton, 1986 อ้างถึงใน กาญจนา วังสุนทร, 2538) จากจุดเด่นดังกล่าวของวิธีแมนเทล-แฮนส์เซลเดิม จึงได้รับความนิยมและการยอมรับในการใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบสองค่า หน่วยงานบริการทดสอบทางการศึกษาแห่งสหรัฐอเมริกา (Educational Testing Service: ETS) ได้แนะนำให้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซลและเป็นวิธีมาตรฐานที่ใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในโครงการสำคัญ ๆ ของหน่วยงาน วิธีการแมนเทล-แฮนส์เซลเดิม จึงถูกปรับขยายเป็น วิธีการแมนเทล-แฮนส์เซลทั่วไป (Generalized Mantel-Haenszel : GMH) ที่สามารถนำมาใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ลำดับชั้นของคะแนนผลการตอบมีมากกว่า 2 ลำดับชั้น และลำดับชั้นการตอบไม่จำเป็นต้องมีการจัดเรียง โดยปรับตารางการถ้อยของคะแนนผลการตอบข้อสอบแบบ $2 \times 2 \times k$ เป็นแบบ $2 \times T \times k$ โดยที่ T เป็นจำนวนลำดับชั้นการตอบ และปรับสถิติทดสอบที่ใช้ในการทดสอบนัยสำคัญจาก MH_{χ^2} ที่มี $df = 1$ เป็น GMH_{χ^2} ที่มี $df = T-1$ เพื่อให้ใช้ได้กับข้อสอบที่มีการให้คะแนนแบบหลายค่า ชวิกค์ และคณะ (Zwick et.al, 1993) ได้ศึกษาประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GMH และวิธี Mantel โดยศึกษากับข้อมูลที่จำลองขึ้นพบว่าในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนหลายค่าด้วยวิธี GMH มีประสิทธิภาพมากกว่าวิธี Mantel เมื่อคะแนนผลการตอบของกลุ่มผู้สอบมีการกระจายมาก และ วิธีการทั้งสองมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันได้เช่นเดียวกับเมื่อตรวจสอบกับข้อสอบที่มีการให้คะแนนแบบสองค่า แต่วิธีนี้มีข้อจำกัดที่สำคัญ คือ การใช้คะแนนที่สังเกตได้แทนความสามารถของกลุ่มผู้สอบทำให้ผลการตรวจสอบพบข้อสอบที่ทำหน้าที่ต่างกันไม่คงที่ เมื่อกลุ่มผู้สอบแปรเปลี่ยนไป แต่จากการศึกษาของ มิลล์เลอร์และสเปร์ย์ (Miller &

Spray, 1993) ที่ศึกษาเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่าด้วยวิธี LDFA กับวิธี GMH โดยใช้ข้อมูลจากการสอบการปฏิบัติวิชาคณิตศาสตร์ พบว่าวิธี GMH ขาดอำนาจในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบไม่สมมาตร ซึ่งสอดคล้องกับการศึกษาของเวลช์และฮูเวอร์ (Miller & Spray, 1993 : 108) ที่พบว่า แม้แต่ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบสมมาตร วิธี GMH ก็มีอำนาจในการตรวจสอบต่ำด้วย และเสนอแนะว่า วิธี GMH อาจจะไม่ได้อันเป็นการทางเลือกที่ดีที่สุด ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่คะแนนผลการตอบเป็นแบบหลายค่า จากการศึกษาของ อุตทาโรและมิลล์ซาป (Uttaro & Millsap, 1994 : 24-25) ได้ให้ความเห็นว่า การศึกษาความถูกต้องของวิธีการแมนเทล-แฮนส์เซลยังต้องศึกษาให้กว้างขวางกว่านี้ เนื่องจากการศึกษาที่ผ่านมาส่วนใหญ่ เป็นการศึกษาจากสถานการณ์จำลองภายใต้เงื่อนไขต่าง ๆ เพื่อพิจารณาความถูกต้องในการตรวจสอบและความผิดพลาดประเภทที่ 1 ว่าอยู่ในระดับที่ยอมรับได้หรือไม่ โดยยังขาดการศึกษาจากข้อมูลที่เป็นสภาพจริงหรือข้อมูลเชิงประจักษ์ และจากการศึกษางานวิจัยที่ศึกษาเกี่ยวกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบสองค่าระหว่างวิธีแมนเทล-แฮนส์เซลกับวิธีที่ใช้ทฤษฎีการตอบข้อสอบ พบว่า ทั้งสองวิธีมีความสอดคล้องกันสูงในการตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน ผู้วิจัยจึงสนใจศึกษา วิธีแมนเทล-แฮนส์เซลรูปแบบทั่วไป (Generalized Mantel-Haenszel : GMH) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่า จากข้อมูลเชิงประจักษ์

วิธีการซิปเทสต์แบบหลายค่า (Polytomous SIBTEST) เป็นวิธีที่ปรับขยายจากวิธี SIBTEST เดิม มีแนวคิดบนพื้นฐานของทฤษฎีการตอบข้อสอบแบบหลายมิติ (Multidimensional IRT) โดยมีข้อตกลงว่ามีมิติของการวัด 2 มิติ มีรูปแบบเป็นนินพาราเมตริก ซึ่งต่างจากวิธี IRT ทั่ว ๆ ไป เนื่องจากไม่ต้องใช้ค่าประมาณความสามารถแฝง วิธีการนี้สามารถใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ทั้งข้อสอบที่เป็นการให้คะแนนแบบสองค่าและคะแนนแบบหลายค่า สามารถตรวจสอบการทำหน้าที่ต่างกันข้อสอบได้ทั้งเป็นรายข้อ (single item) และเป็นกลุ่มข้อสอบ (item-bundle) และสามารถวิเคราะห์ได้ทั้งแบบทดสอบที่เป็นมิติเดียว (unidimensional test) และแบบทดสอบที่เป็นหลายมิติ (multidimensional test) (Stout, Li and Nandakumar, 1997 : 195) เป็นวิธีที่ใช้คะแนนรวมจากแบบทดสอบเป็นตัวแทนความสามารถของผู้สอบและใช้สถิติทดสอบนัยสำคัญ (Narayanan and Swaminathan, 1994 : 315) วิธีการนี้ได้รับความนิยมเช่นเดียวกับวิธีแมนเทล-แฮนส์เซล เนื่องจากเป็นวิธีที่ประหยัด มีการคำนวณที่ง่าย ไม่ซับซ้อน ไม่จำเป็นต้องใช้กลุ่มตัวอย่างขนาดใหญ่ จากการศึกษาของนารายานันและสวามินาธาน (Narayanan and Swaminathan, 1994) ที่เปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบไม่สมมาตร ด้วยวิธี MH และวิธี SIBTEST พบว่า ทั้งสองวิธีมีความถูกต้องในการตรวจสอบเท่าเทียมกัน แต่วิธี SIBTEST มีความผิดพลาดประเภทที่ 1 สูงกว่า ส่วนการศึกษาของรูสโซและสตาท์ (Roussos and Stout, 1996) ที่ศึกษาผลกระทบของกลุ่มตัวอย่างที่มีต่อความผิดพลาดประเภทที่ 1 ของวิธี MH และวิธี SIBTEST พบว่า วิธี MH มีความผิดพลาดประเภทที่ 1 สูงกว่า จิตติมา วรรณศรี (2539) ได้เปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบไม่สมมาตร ด้วยวิธี MH และวิธี SIBTEST พบว่า ทั้งสองวิธีมีความถูกต้องในการตรวจสอบเท่าเทียมกัน แต่วิธี SIBTEST มีความผิดพลาดประเภทที่ 1 สูงกว่า จากการศึกษาดังกล่าวข้างต้นเป็นการศึกษากับข้อมูลที่เป็นแบบสองค่า ผู้วิจัยจึงสนใจที่จะศึกษาวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทั้งวิธี GMH และวิธี Polytomous SIBTEST โดยนำวิธี GPCM ซึ่งเป็นวิธีการตามทฤษฎีการตอบข้อสอบ (IRT) มาเป็นเกณฑ์ในการตัดสินใจว่า ข้อสอบทำหน้าที่ต่างกัน โดยใช้ข้อสอบวิชาคณิตศาสตร์ที่มีลักษณะเป็นข้อสอบแบบถูกผิดหลายตัวเลือกที่มีการให้คะแนนแบบหลายค่า

จากการศึกษางานวิจัย ที่เปรียบเทียบวิธีการในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ผ่านมา พบว่า ปัจจัยสำคัญที่ส่งผลต่อประสิทธิภาพในการตรวจสอบ คือ ขนาดของกลุ่มตัวอย่าง เช่น จากการศึกษาของ สวามินาทานและโรเจอร์ (Swaminathan & Rogers, 1990) ใช้กลุ่มตัวอย่างขนาด 250 และ 500 คน ตรวจสอบการทำหน้าที่ต่างกันทั้งแบบสม่ำเสมอและแบบไม่สม่ำเสมอ ด้วยวิธี MH และวิธีการถดถอยโลจิสติก พบว่า เมื่อเพิ่มขนาดของกลุ่มตัวอย่างจะมีผลทำให้อำนาจการตรวจสอบของวิธีการถดถอยโลจิสติกเพิ่มมากขึ้นเกือบทุกเงื่อนไข ส่วนวิธี MH อำนาจการตรวจสอบเพิ่มมากขึ้นเฉพาะกรณีที่เป็นแบบสม่ำเสมอและความผิดพลาดประเภทที่ 1 ของวิธี MH ลดลงเกือบทุกเงื่อนไข มาเซอร์และคณะ (Mazor et al. 1992) ได้ศึกษาวิธี MH จากข้อมูลจำลองโดยใช้กลุ่มตัวอย่างขนาด 100, 200, 500, 1,000 และ 2,000 คน รวมทั้ง นารายานันและสวามินาทาน (Narayanan & Swaminathan, 1994) ศึกษาผลของการตรวจสอบด้วยวิธี MH และวิธี SIBTEST กับกลุ่มตัวอย่าง 3 ขนาด คือ 300, 500 และ 1,000 คน ได้ผลการศึกษาที่สอดคล้องกัน คือเมื่อเพิ่มขนาดกลุ่มตัวอย่างให้มากขึ้น ความถูกต้องในการตรวจสอบจะสูงขึ้น ฮิลล์ (Hill, 1990 cited in Mazor et al. 1992) กล่าวว่า สำหรับวิธี MH ควรใช้กลุ่มตัวอย่างขนาดระหว่าง 100 ถึง 300 คน สำหรับกลุ่มใดกลุ่มหนึ่งหรือทั้งสองกลุ่ม ส่วนมาเซอร์และคณะ (Mazor et al. 1992) เสนอว่า การใช้กลุ่มตัวอย่างขนาด 200 คน ก็เพียงพอแล้วและไม่ควรน้อยกว่านี้ ซึ่งนารายานันและสวามินาทาน (Narayanan & Swaminathan, 1994) เสนอว่า สำหรับวิธี MH และวิธี SIBTEST ขนาดของกลุ่มตัวอย่างกลุ่มละ 300 คน ก็เพียงพอที่จะตรวจสอบอย่างมีประสิทธิภาพ ส่วนเฟรนช์และมิลเลอร์ (French & Miller, 1996) ได้ศึกษาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติกที่ข้อสอบมีการให้คะแนนแบบหลายค่า พบว่า เมื่อกลุ่มตัวอย่างมีขนาดเล็กลงอำนาจในการตรวจสอบจะลดลง ต่อมาจิตติมา วรณศรี (2539) ได้ศึกษา พบว่า เมื่อขนาดกลุ่มตัวอย่าง 200 คนและ 600 คน ทั้งวิธี MH และวิธี SIBTEST สามารถตรวจสอบได้ถูกต้อง 50% แต่ถ้าขนาดของกลุ่มตัวอย่างเพิ่มขึ้นเป็น 1,000 คน สามารถตรวจสอบได้ถูกต้อง 100% นอกจากนั้น ปัจจัยที่ส่งผลต่อประสิทธิภาพในการตรวจสอบอีกอย่างหนึ่ง คือ ความยาวของแบบทดสอบ ทั้งนี้ เพราะความยาวของแบบทดสอบจะมีผลกระทบต่อความถูกต้องในการจับคู่เปรียบเทียบกลุ่มผู้สอบ โดยเฉพาะวิธี GMH และวิธี Polytomous SIBTEST ที่ใช้คะแนนรวมจากการทดสอบแทนคุณลักษณะภายในหรือความสามารถของผู้สอบ แบบทดสอบที่มีความยาวมากกว่าย่อมส่งผลให้มีความน่าเชื่อถือมากกว่า เนื่องจากการจับคู่เปรียบเทียบระหว่างผู้สอบมีความถูกต้องมากขึ้น เช่น ผลการศึกษาของสวามินาทานและโรเจอร์ (Swaminathan & Rogers, 1990) พบว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทั้งแบบสม่ำเสมอและแบบไม่สม่ำเสมอ เมื่อใช้แบบทดสอบที่มีความยาวมากขึ้น มีผลทำให้ความถูกต้องในการตรวจสอบของวิธี MH และวิธีการถดถอยโลจิสติก เพิ่มขึ้น ยกเว้น ในกรณีการทำหน้าที่ต่างกันของข้อสอบแบบไม่สม่ำเสมอของวิธี MH ต่อมาทั้งสองคนได้ทำการศึกษาใหม่ (Rogers & Swaminathan, 1993) พบว่า ผลที่ได้ขัดแย้งกับครั้งแรก คือ ความยาวของแบบทดสอบไม่มีผลต่ออำนาจการทดสอบของวิธี MH และวิธีการถดถอยโลจิสติก ยกเว้น ในกรณีการทำหน้าที่ต่างกันของข้อสอบแบบไม่สม่ำเสมอของวิธีการถดถอยโลจิสติก ส่วนการศึกษาของ กาญจนา วจินสุนทร (2538) พบว่า ความยาวของแบบทดสอบไม่มีผลกระทบต่อความถูกต้องในการตรวจสอบของวิธี MH และวิธี SIBTEST ซึ่งขัดแย้งกับผลการศึกษาของ จิตติมา วรณศรี (2539) ที่พบว่าแบบทดสอบที่มีขนาด 60 ข้อ จะมีผลทำให้วิธี MH และวิธี SIBTEST มีประสิทธิภาพในการตรวจสอบดีที่สุด ส่วนอุตทาโรและมิลล์ซาป (Uttaro & Millsap, 1994) ได้ศึกษาพบว่า เมื่อใช้แบบทดสอบที่มีความยาวมากขึ้นแล้ว ความผิดพลาดประเภทที่ 1 ของวิธี MH จะมีค่าลดลง

จากผลการศึกษาดังกล่าวข้างต้น ผู้วิจัยจึงมีความสนใจที่จะเปรียบเทียบประสิทธิภาพของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนแบบหลายค่าระหว่างวิธี GMH และวิธี Polytomous SIBTEST โดยใช้คะแนนที่ได้จากการตอบข้อสอบวิชาคณิตศาสตร์ เรื่อง วงกลม ชั้นมัธยมศึกษาปีที่ 3 ซึ่งเป็นข้อมูลเชิงประจักษ์ ตามปัจจัยด้านขนาดของกลุ่มตัวอย่างและความยาวของแบบทดสอบ เพื่อเป็นแนวทางในการตัดสินใจเลือกใช้วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้อย่างเหมาะสมต่อไป

ความมุ่งหมายของการวิจัย

การวิจัยครั้งนี้มีจุดมุ่งหมายสำคัญ เพื่อศึกษาวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่เหมาะสมกับข้อมูลผลการตอบข้อสอบที่ให้คะแนนหลายค่า โดยกำหนดเป็นจุดมุ่งหมายเฉพาะในการศึกษาวิธี GMH และวิธี Polytomous SIBTEST ด้านต่าง ๆ ดังนี้

1. เปรียบเทียบความถูกต้องและความผิดพลาดของการตรวจสอบการทำหน้าที่ต่างกันของวิธี GMH ระหว่างกลุ่มตัวอย่างขนาดต่างกัน
2. เปรียบเทียบความถูกต้องและความผิดพลาดของการตรวจสอบการทำหน้าที่ต่างกันของวิธี Polytomous SIBTEST ระหว่างกลุ่มตัวอย่างขนาดต่างกัน
3. เปรียบเทียบความถูกต้องและความผิดพลาดของการตรวจสอบการทำหน้าที่ต่างกัน ระหว่างวิธี GMH และวิธี Polytomous SIBTEST เมื่อใช้กลุ่มตัวอย่างขนาดเท่ากัน
4. เพื่อศึกษาความถูกต้องและความผิดพลาดของการตรวจสอบการทำหน้าที่ต่างกัน ระหว่างวิธี GMH และวิธี Polytomous SIBTEST เมื่อความยาวของแบบทดสอบมีขนาดเท่ากัน

ความสำคัญของการวิจัย

การวิจัยครั้งนี้ เป็นการศึกษาค้นคว้าเกี่ยวกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ใช้ข้อมูลผลการตอบข้อสอบที่มีการให้คะแนนแบบหลายค่า ซึ่งเป็นการขยายขอบเขตของการศึกษาค้นคว้าเกี่ยวกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่ใช้ข้อมูลผลการตอบข้อสอบที่มีการให้คะแนนแบบสองค่า เนื่องจากยังไม่พบการเผยแพร่ผลการศึกษาเกี่ยวกับ วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ใช้ข้อมูลผลการตอบข้อสอบที่มีการให้คะแนนแบบหลายค่าดังกล่าวมาก่อน และผลจากการศึกษานี้จะเป็นแนวทางในการตัดสินใจเลือกใช้วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้อย่างเหมาะสมกับขนาดกลุ่มตัวอย่างและความยาวของแบบทดสอบ และเป็นแนวทางในการปรับปรุงข้อสอบให้มีความยุติธรรมกับกลุ่มผู้สอบ ซึ่งเป็นทางเลือกอีกทางหนึ่งในการพัฒนาคุณภาพของเครื่องมือวัดต่อไป

ขอบเขตของการวิจัย

1. ข้อมูลที่ใช้ในการศึกษา เป็นข้อมูลจริงที่ได้จากการตอบข้อสอบวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 3 เรื่องวงกลม ที่ผู้วิจัยสร้างขึ้น โดยมีวัตถุประสงค์ให้ข้อสอบส่วนหนึ่งของแบบทดสอบ แสดงการทำหน้าที่ต่างกันที่เอื้อต่อเพศชาย กลุ่มตัวอย่างที่ใช้ในการเก็บรวบรวมข้อมูลเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1 ปีการศึกษา 2545 ของโรงเรียนสังกัดกรมสามัญศึกษาในจังหวัดพระนครศรีอยุธยา

2. ตัวแปรที่ศึกษา ประกอบด้วย

2.1 ตัวแปรอิสระ ได้แก่

2.1.1 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 2 วิธี ได้แก่

2.1.1.1 วิธี GMH (Generalized Mantel-Haenszel)

2.1.1.2 วิธี Polytomous SIBTEST

2.1.2 ขนาดของกลุ่มตัวอย่างที่ศึกษามี 3 ขนาด คือ

2.1.2.1 กลุ่มอ้างอิง ต่อ กลุ่มสนใจ เท่ากับ 250 คน ต่อ 250 คน

2.1.2.2 กลุ่มอ้างอิง ต่อ กลุ่มสนใจ เท่ากับ 500 คน ต่อ 500 คน

2.1.2.3 กลุ่มอ้างอิง ต่อ กลุ่มสนใจ เท่ากับ 1,000 คน ต่อ 1,000 คน

2.2 ตัวแปรตาม คือ ประสิทธิภาพในการตรวจสอบ โดยพิจารณาจาก

2.2.1 ความถูกต้องของการตรวจสอบ

2.2.2 ความผิดพลาดของการตรวจสอบ

2.3 ตัวแปรควบคุม คือ ความยาวของแบบทดสอบมี 3 ขนาด ได้แก่ แบบทดสอบที่ประกอบด้วยข้อสอบ 40 ข้อ, 30 ข้อ และ 20 ข้อ

3. โมเดลที่ใช้ในการศึกษาครั้งนี้ คือ โมเดลของทฤษฎีการตอบข้อสอบแบบโลจิสติก 2 PLM ที่การให้คะแนนเป็นแบบหลายค่า (Polytomous IRT Model)

4. เกณฑ์ที่ใช้ในการวิเคราะห์ว่าข้อสอบทำหน้าที่ต่างกัน

4.1 วิธี GMH ได้แก่ ค่า $\hat{\alpha}_{GMH} \neq 1$ และทดสอบนัยสำคัญด้วยสถิติ $GMH\chi^2$ ที่ระดับ .05

4.2 วิธี Polytomous SIBTEST ได้แก่ ค่า $\hat{\beta}^* > 0$ และทดสอบนัยสำคัญด้วยสถิติ Z ที่ระดับ .05

4.3 วิธี GPCM ได้แก่ การทดสอบนัยสำคัญของค่าพารามิเตอร์ของข้อสอบด้วยสถิติ χ^2 ที่ระดับ .05

นิยามศัพท์เฉพาะ

1. ข้อสอบที่มีการให้คะแนนแบบหลายค่า (Polytomously Scored Items) หมายถึง ข้อสอบที่มีการกำหนดลำดับขั้นของการให้คะแนนผลการตอบข้อสอบไว้เป็นจำนวนเต็มหลายระดับคะแนน ซึ่งในการวิจัยนี้กำหนดไว้ 6 ระดับคะแนน คือ 0, 1, 2, 3, 4 และ 5 คะแนน โดยใช้ข้อสอบเลือกตอบแบบถูกผิด (Multiple True-False : MTF) ชนิด 5 ตัวเลือก ในวิชาคณิตศาสตร์ เรื่อง วงกลม ชั้นมัธยมศึกษาปีที่ 3

2. กลุ่มอ้างอิง (Reference group) หมายถึง กลุ่มผู้สอบที่คาดว่าจะได้ประโยชน์จากการตอบข้อสอบที่ทำหน้าที่ต่างกัน คือ เป็นกลุ่มที่มีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องสูงกว่าผู้สอบอีกกลุ่มหนึ่งทั้ง ๆ ที่มีความสามารถเท่ากัน ในการวิจัยครั้งนี้กลุ่มอ้างอิง คือ เพศชาย

3. กลุ่มสนใจ (Focal group) หมายถึง กลุ่มผู้สอบที่คาดว่าจะเสียประโยชน์จากการตอบข้อสอบที่ทำหน้าที่ต่างกัน คือ เป็นกลุ่มที่มีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องต่ำกว่าผู้สอบอีกกลุ่มหนึ่ง ทั้ง ๆ ที่มีความสามารถเท่ากัน ในการวิจัยครั้งนี้กลุ่มสนใจ คือ เพศหญิง

4. การทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning : DIF) หมายถึง การที่ข้อสอบทำให้ผู้สอบซึ่งเป็นสมาชิกของกลุ่มอ้างอิง (เพศชาย) และผู้สอบที่เป็นสมาชิกของกลุ่มสนใจ (เพศหญิง) ที่มีความสามารถในสิ่งที่ต้องการวัดเท่ากัน มีโอกาสตอบข้อสอบข้อนั้นได้ถูกต้องไม่เท่ากัน

5. ขนาดของกลุ่มตัวอย่าง หมายถึง จำนวนผู้สอบของกลุ่มอ้างอิงต่อกลุ่มสนใจ ซึ่งในการวิจัยนี้แบ่งกลุ่มตัวอย่างเป็น 3 ขนาด คือ 250 : 250, 500 : 500 และ 1,000 : 1,000

6. ความยาวของแบบทดสอบ หมายถึง จำนวนของข้อสอบในแบบทดสอบ ในการวิจัยนี้ศึกษาความยาวของแบบทดสอบ 3 ขนาด คือ ความยาว 40 ข้อ, 30 ข้อ และ 20 ข้อ

7. วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (Detecting DIF Procedures) หมายถึง วิธีการวิเคราะห์ทางสถิติที่ใช้เพื่อบ่งบอกว่าการทำหน้าที่ต่างกันของข้อสอบระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจ ซึ่งวิธีการในงานวิจัยนี้เป็นวิธีการตรวจสอบที่ใช้กับข้อสอบที่มีการให้คะแนนแบบหลายค่า ได้แก่

7.1 วิธี GMH หมายถึง วิธีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ตามแนวคิดของวิธี Mantel-Haenszel เดิมมาทำการปรับขยายใช้กับข้อสอบที่มีการให้คะแนนแบบหลายค่า โดยการวิเคราะห์ความแตกต่างของสัดส่วนการตอบข้อสอบระหว่างผู้สอบกลุ่มย่อย 2 กลุ่มที่มีความสามารถในระดับเดียวกัน โดยใช้ค่า $\hat{\alpha}_{GMH}$

7.2 วิธี Polytomous SIBTEST หมายถึง วิธีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ตามแนวคิดของวิธี SIBTEST เดิมมาทำการปรับปรุงใหม่เพื่อใช้กับข้อสอบที่มีการให้คะแนนแบบหลายค่า โดยพิจารณาจากความแตกต่างของคะแนนจริงระหว่างผู้สอบกลุ่มย่อย 2 กลุ่มที่มีความสามารถระดับเดียวกัน โดยใช้ค่า $\hat{\beta}^*$

7.3 วิธี GPCM หมายถึง วิธีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ตามแนวทฤษฎีการตอบข้อสอบ (IRT) ที่ใช้กับข้อสอบที่มีการให้คะแนนแบบหลายค่า โดยพิจารณาจากความแตกต่างของฟังก์ชันการตอบข้อสอบของผู้สอบที่มีความสามารถเท่ากัน แต่อยู่ต่างกลุ่มกัน ในการตอบข้อสอบชุดเดียวกัน โดยใช้ค่า χ^2

8. เกณฑ์การตัดสินว่าข้อสอบทำหน้าที่ต่างกัน หมายถึง ค่าดัชนีที่ใช้ตัดสินว่าข้อสอบทำหน้าที่ต่างกัน ซึ่งเกณฑ์การตัดสินของแต่ละวิธีการ เป็นดังนี้

8.1 เกณฑ์การตัดสินตามวิธี GMH หมายถึง ค่า $\hat{\alpha}_{GMH}$ แตกต่างจาก 1 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

8.2 เกณฑ์การตัดสินตามวิธี Polytomous SIBTEST หมายถึง ค่า $\hat{\beta}^*$ แตกต่างจาก 0 อย่างมีนัยสำคัญทางสถิติ ที่ระดับ .05

8.3 เกณฑ์การตัดสินตามวิธี GPCM หมายถึง ค่าความแตกต่างของค่าพารามิเตอร์ข้อสอบระหว่างผู้สอบย่อย 2 กลุ่ม อย่างมีนัยสำคัญด้วยสถิติ χ^2 ที่ระดับ .05

9. ประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ หมายถึง ความถูกต้องของการระบุการทำหน้าที่ต่างกันของข้อสอบ จากการตรวจสอบด้วยวิธี GMH และวิธี Polytomous SIBTEST ซึ่งพิจารณาได้จากความถูกต้องของการตรวจสอบและความผิดพลาดของการตรวจสอบ

10. ความถูกต้องของการตรวจสอบ หมายถึง อัตราส่วนหรือร้อยละของจำนวนข้อสอบที่ตรวจสอบพบว่าทำหน้าที่ต่างกันได้อย่างถูกต้อง โดยคำนวณจากค่าอัตราส่วนของจำนวนข้อสอบที่ตรวจสอบพบว่าทำหน้าที่ต่างกันได้อย่างถูกต้องต่อจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งหมดในแบบทดสอบ ซึ่งในการวิจัยนี้ จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งหมดในแบบทดสอบ ผู้วิจัยใช้วิธี GPCM เป็นเกณฑ์มาตรฐานตัดสิน

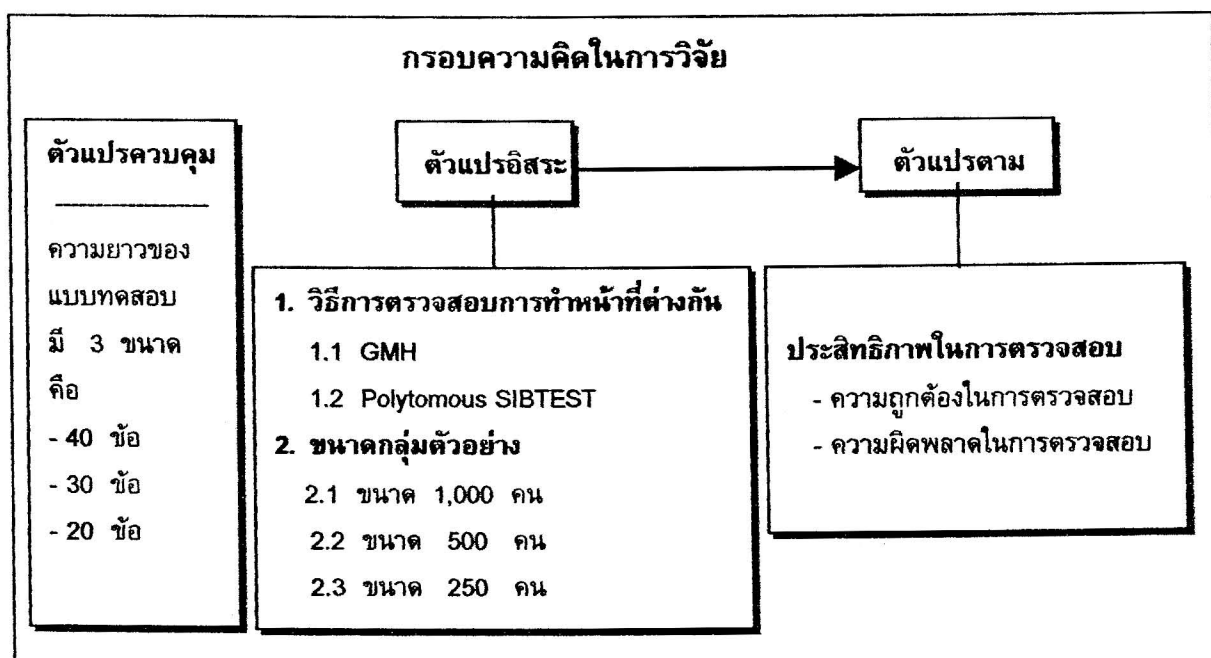
11. ความผิดพลาดของการตรวจสอบ หมายถึง อัตราส่วนหรือร้อยละของจำนวนข้อสอบที่ระบุผิดพลาด ซึ่งมี 2 ประเภท คือ

11.1 ความผิดพลาดประเภทที่ 1 (Type I Error Rate) คือ การระบุผิดพลาดว่าข้อสอบทำหน้าที่ต่างกัน (False Positive : FP) ทั้งที่ความเป็นจริงข้อสอบทำหน้าที่ไม่ต่างกัน ซึ่งคำนวณได้จากค่าอัตราส่วนของจำนวนข้อสอบที่ระบุผิดพลาดว่าทำหน้าที่ต่างกันต่อจำนวนข้อสอบที่ทำหน้าที่ไม่ต่างกันทั้งหมดในแบบทดสอบ ซึ่งในการวิจัยครั้งนี้พิจารณาเฉพาะ ความผิดพลาดประเภทที่ 1 เท่านั้น

11.2 ความผิดพลาดประเภทที่ 2 (Type II Error Rate) คือ การระบุผิดพลาดว่าข้อสอบทำหน้าที่ไม่ต่างกัน (False Negative : FN) ทั้งที่ความเป็นจริงข้อสอบทำหน้าที่ต่างกัน ซึ่งคำนวณได้จากค่าอัตราส่วนของจำนวนข้อสอบที่ระบุผิดพลาดว่าทำหน้าที่ไม่ต่างกันต่อจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งหมดในแบบทดสอบ

กรอบความคิดในการวิจัย

ประเด็นการศึกษา : ในการเปรียบเทียบประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยอาศัยข้อมูลจากการตอบข้อสอบที่มีการให้คะแนนแบบหลายค่า นั้น มีกรอบความคิด ดังนี้



ภาพประกอบ 1 กรอบความคิดในการวิจัย

บทที่ 2

เอกสารงานวิจัยที่เกี่ยวข้อง

ในการศึกษาครั้งนี้ ผู้วิจัยนำเสนอแนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง เรียงลำดับดังนี้

1. แนวคิดเกี่ยวกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ
2. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี GMH
3. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี Polytomous SIBTEST
4. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี GPCM
5. งานวิจัยที่เกี่ยวข้องกับวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการตรวจให้คะแนนแบบหลายค่า

1. แนวคิดเกี่ยวกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

การศึกษาเรื่องความยุติธรรมของข้อสอบ ในกรณีนี้ข้อสอบทำให้ผู้สอบระหว่างกลุ่มย่อยเกิดการได้เปรียบเสียเปรียบกัน เดิมใช้คำว่า “ความลำเอียงของข้อสอบ” (Item Bias) ซึ่งเป็นภาษาที่ใช้กันในทางสังคม และมีความหมายไปในทางลบ ภายหลังจากนักวิจัยได้เปลี่ยนไปใช้คำใหม่ว่า “การทำหน้าที่ต่างกันของข้อสอบ” (Differential Item Functioning : DIF) เนื่องจากเห็นว่าเป็นคำที่มีความหมายกลาง ๆ จึงมีความเหมาะสมเชิงวิชาการมากกว่า อย่างไรก็ตาม คำสองคำนี้มีจุดเน้นที่แตกต่างกัน คำว่า “ความลำเอียงของข้อสอบ” เน้นที่อิทธิพลที่สังเกตได้ของกลุ่มผู้สอบย่อยที่มุ่งศึกษา ส่วนคำว่า “ข้อสอบทำหน้าที่ต่างกัน” นั้น เน้นที่คุณลักษณะทางสถิติของข้อสอบที่ตรวจสอบได้ด้วยวิธีการวิเคราะห์ทางสถิติ ซึ่งเป็นส่วนประกอบหนึ่งของสิ่งที่แสดงถึงความลำเอียงของข้อสอบ วิธีการทางสถิติที่ใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเป็นเงื่อนไขที่จำเป็น (necessary condition) ในการตัดสินความลำเอียงของข้อสอบ แต่ไม่ใช่เงื่อนไขที่เพียงพอ (not sufficient condition) ในการตัดสินความลำเอียงของข้อสอบ เนื่องจากถ้าใช้วิธีการทางสถิติตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเพียงอย่างเดียวแล้ว ผลการตรวจสอบพบว่าข้อสอบทำหน้าที่ต่างกันนั้นยังสรุปไม่ได้ว่าข้อสอบลำเอียงหรือไม่ ยังต้องใช้ผู้เชี่ยวชาญพิจารณาเนื้อหาสาระของข้อสอบและจุดมุ่งหมายในการวัดของข้อสอบ ที่เรียกว่า “วิธีการตัดสินข้อสอบ” (judgemental method) (Camilli & Shepard, 1994 : 135)

มีผู้ให้ความหมายของคำว่า “การทำหน้าที่ต่างกันของข้อสอบ” ไว้ดังนี้

ดูรานส์และคูลิค (Dorans and Kulick, 1986 : 355) กล่าวว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง โอกาสในการตอบข้อสอบได้ถูกต้องของผู้สอบกลุ่มหนึ่งมีค่าต่ำกว่าหรือสูงกว่า ผู้สอบอีกกลุ่มหนึ่งที่มีระดับความสามารถเดียวกัน

มิลล์แซฟและเอฟเวอร์สัน (Millsap and Everson, 1993 : 298) กล่าวว่า ข้อสอบที่ทำหน้าที่ต่างกัน หมายถึง ความแตกต่างในการทำหน้าที่ของข้อสอบระหว่างผู้สอบกลุ่มย่อยที่ได้ถูกจับคู่ตามความสามารถหรือคุณลักษณะ (attribute) ที่ข้อสอบข้อนั้นวัด

คิมและคณะ (Kim et al. 1994 : 217) กล่าวว่า ข้อสอบทำหน้าที่ต่างกันเมื่อฟังก์ชันการตอบข้อสอบจากผู้สอบกลุ่มย่อยที่อยู่ต่างกลุ่มกัน มีค่าต่างกัน

โปเทนซ่าและดอรานส์ (Potenza and Dorans. 1995 : 23) กล่าวว่า ถ้าข้อสอบทำหน้าที่ต่างกัน จะทำให้ผลการตอบข้อสอบระหว่างผู้สอบกลุ่มย่อย ที่นำมาเปรียบเทียบกันแตกต่างกัน

มาเซอร์และคลอสเซอร์ (Mazor and Clauser. 1995 : 131) กล่าวว่า ข้อสอบทำหน้าที่ต่างกัน เมื่อผู้สอบที่มีความสามารถระดับเดียวกัน แต่เป็นสมาชิกของกลุ่มย่อยต่างกัน มีโอกาสในการตอบข้อสอบถูกต้องแตกต่างกัน

โดยสรุป “การทำหน้าที่ต่างกันของข้อสอบ หมายถึง การที่ข้อสอบให้โอกาสในการตอบข้อสอบแก่ผู้สอบที่มีความสามารถเท่ากันในสิ่งที่ต้องการวัดไม่เท่ากัน เนื่องจากผู้สอบอยู่ในกลุ่มประชากรย่อยที่แตกต่างกัน”

ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบดำเนินการโดยการเปรียบเทียบผลการตอบข้อสอบระหว่างผู้สอบ 2 กลุ่มที่มีความสามารถระดับเดียวกัน โดยกำหนดให้ผู้สอบกลุ่มหนึ่ง เป็น “กลุ่มอ้างอิง” (Reference group : R) ซึ่งเป็นกลุ่มที่คาดว่าจะได้ประโยชน์ในการตอบข้อสอบ คือ มีโอกาสในการตอบข้อสอบถูกมากกว่าอีกกลุ่มหนึ่ง ส่วนอีกกลุ่มหนึ่ง เป็น “กลุ่มเปรียบเทียบหรือกลุ่มสนใจ” (Focal group : F) ซึ่งเป็นกลุ่มที่คาดว่าจะเสียประโยชน์ในการตอบข้อสอบ คือ มีโอกาสในการตอบข้อสอบได้ถูกต้องน้อยกว่าผู้สอบอีกกลุ่มหนึ่ง สำหรับเกณฑ์ที่ใช้ในการจำแนกผู้สอบเป็นกลุ่มสนใจและกลุ่มอ้างอิงมีหลายลักษณะ เช่น เพศ สีผิว เชื้อชาติ ภาษา วัฒนธรรม ภูมิฐานะ เป็นต้น (Holland and Wainer. 1993 : Xiii)

เมลเลนเบอร์ก (Mellenbergh. 1989 : 127-143) ได้จำแนกประเภทของการทำหน้าที่ต่างกันของข้อสอบ ออกเป็น 2 ประเภท คือ

1. การทำหน้าที่ต่างกันของข้อสอบแบบสม่ำเสมอ (Uniform DIF) เกิดขึ้นเมื่อผู้สอบกลุ่มหนึ่งมีโอกาสในการตอบข้อสอบถูกมากกว่าผู้สอบอีกกลุ่มหนึ่งในทุกระดับความสามารถ เมื่อพิจารณาโค้งคุณลักษณะข้อสอบ (Item Characteristic Curves : ICC) ระหว่างผู้สอบกลุ่มย่อย 2 กลุ่มจะขนานกันหรือไม่มีปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้สอบ และ การเป็นสมาชิกของกลุ่ม (group membership)

2. การทำหน้าที่ต่างกันของข้อสอบแบบไม่สม่ำเสมอ (Nonuniform DIF) เกิดขึ้นเมื่อโอกาสในการตอบข้อสอบถูกของผู้สอบระหว่างกลุ่มย่อย 2 กลุ่มไม่สม่ำเสมอ เมื่อพิจารณาแต่ละระดับความสามารถ หรือ เมื่อพิจารณาโค้งคุณลักษณะข้อสอบระหว่างผู้สอบกลุ่มย่อย 2 กลุ่มจะไม่ขนานกันหรือมีปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้สอบ และ การเป็นสมาชิกของกลุ่ม

วิธีการในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบมีหลายวิธี สามารถจำแนกได้หลายลักษณะ ขึ้นอยู่กับเกณฑ์ที่ใช้ในการจำแนก เช่น **การใช้เกณฑ์การให้คะแนน** แบ่งได้เป็น 2 กลุ่มวิธี คือ กลุ่มวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบ 2 ค่า (dichotomous DIF procedures) กลุ่มนี้ข้อสอบที่ตรวจสอบการทำหน้าที่ต่างกัน มีการให้คะแนนเป็นแบบ 0 – 1 เช่น แบบทดสอบเลือกตอบที่ให้คะแนนการตอบถูกเป็น 1 คะแนนและตอบผิดเป็น 0 คะแนน และ กลุ่มวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่า (Polytomous DIF procedures) เช่น ข้อสอบวัดการปฏิบัติ (performance test) ข้อสอบที่ให้สร้างคำตอบเอง (constructed-response items) ไม่ว่าจะเป็นข้อสอบที่วัดการอ่าน (Reading item) หรือการเขียน (Writing item) และข้อสอบปลายเปิด (open-ended item) หรือแบบทดสอบเลือกตอบที่มีการให้คะแนนความรู้บางส่วน เช่น แบบทดสอบเลือกตอบแบบถูกผิด เป็นต้น **การใช้เกณฑ์ที่ยึดทฤษฎีของการวิเคราะห์ข้อมูล** แบ่งเป็น 2 กลุ่มวิธี คือ กลุ่มวิธีที่ยึดทฤษฎี IRT ที่วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบโดยใช้คะแนนที่สังเกตไม่ได้ หรือ ตัวแปรแฝงภายใต้ทฤษฎีการตอบข้อสอบ (Item Response Theory) และ กลุ่มวิธีที่ไม่ใช่ IRT (non IRT) กลุ่มนี้จะวิเคราะห์การทำหน้าที่ต่างกันของ

ข้อสอบโดยใช้คะแนนที่สังเกตได้ภายใต้ทฤษฎีการทดสอบมาตรฐานเดิม (Classical Test Theory) การใช้เกณฑ์ข้อตกลงเบื้องต้นของโมเดล แบ่งเป็น 2 กลุ่มวิธี คือ กลุ่มวิธีที่ยึดรูปแบบพาราเมตริก (Parametric form) การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบมีข้อตกลงเบื้องต้นของโมเดลสำหรับอธิบายความสัมพันธ์ระหว่างคะแนนของข้อสอบและการจับคู่ตัวแปร และ กลุ่มวิธีที่ยึดรูปแบบนพาราเมตริก (Nonparametric form) ซึ่งกลุ่มนี้จะไม่มีข้อตกลงเบื้องต้นดังกล่าว

โปเทนซ่าและดูรานส์ ได้จำแนกวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่า สามารถสรุปรวมเป็นตารางได้ดังนี้ (Potenza and Dorans. 1995 : 25)

ตาราง 1 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่า จำแนกตามทฤษฎีที่ใช้ในการวิเคราะห์และรูปแบบของโมเดล

วิธีการตรวจสอบ DIF	รูปแบบพาราเมตริก	รูปแบบนพาราเมตริก
กลุ่มวิธี non-IRT	- วิธีการถดถอยโลจิสติกแบบหลายค่า (Polytomous LRDIF)	- วิธีแมนเทล MNTL _{P-DIF}
	- วิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก (Logistic Discriminant Function Analysis : LDFA)	- วิธีการมาตรฐานแบบหลายค่า (Polytomous STND)
กลุ่มวิธี IRT	- วิธีการทดสอบอัตราส่วนความน่าจะเป็น (General IRT likelihood ratio)	- วิธี HW1
	- วิธีฟาเชี่ยล เครดิต โมเดล (Patial Credit Model : PCM)	- วิธี HW3
		- วิธีแมนเทล-แฮนส์เซลทั่วไป (Generalized Mantel – Haenszel : GMH)
		- วิธีซิปเทสต์แบบหลายค่า (Polytomous SIBTEST)
		- วิธีฟาเชี่ยล เครดิต โมเดลทั่วไป (Generalized Patial Credit Model : GPCM)

วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนหลายค่าที่ผู้วิจัยสนใจศึกษาครั้งนี้ คือ วิธี GMH และวิธี Polytomous SIBTEST

2. วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี GMH (Generalized Mantel–Haenszel Procedure : GMH)

วิธี GMH เป็นวิธีการที่พัฒนาจากวิธี MH เดิม โดยปรับขยายเพื่อใช้กับข้อสอบที่มีการให้คะแนนแบบหลายค่า (Polytomously Scored Items) วิธี MH เดิม เป็นวิธีที่พัฒนาจากวิธีไค-สแควร์แบบเดิม (traditional χ^2 approach) ซึ่งเสนอโดย แมนเทล-แฮนส์เซล ในปี ค.ศ. 1959 เพื่อใช้ในการทำวิจัยทางการ

แพทช์ ฮอลแลนด์และธาเยอร์ (Holland and Thayer. 1988) เป็นผู้ริเริ่มนำวิธี MH มาประยุกต์ใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของ หน่วยงานบริการทดสอบทางการศึกษา แห่งสหรัฐอเมริกา (ETS) ทำให้วิธี MH เป็นที่ยอมรับและถูกนำมาใช้อย่างกว้างขวาง วิธี MH เป็นวิธีตารางการณั้จรแบบ ไม่คำนวณทวนซ้ำ (noniterative contingency table method) สามารถนำไปใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ง่าย มีขั้นตอนการคำนวณที่สะดวก ไม่สลับซับซ้อน ประหยัด มีการทดสอบทางสถิติแบบ นันพารามेटริก (nonparametric) คือ ไม่จำเป็นต้องใช้โมเดลประมาณค่า

หลักการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยวิธี MH เป็นการเปรียบเทียบผลการตอบ ข้อสอบของผู้สอบ 2 กลุ่มได้แก่ กลุ่มอ้างอิง (reference group) ซึ่งเป็นกลุ่มที่ใช้เป็นมาตรฐานในการเปรียบเทียบ ผลการตอบข้อสอบกับผู้สอบอีกกลุ่มหนึ่ง ถ้าข้อสอบที่นำมาศึกษาทำหน้าที่ต่างกันแล้ว คาดว่าผู้สอบ กลุ่มนี้จะได้ประโยชน์ในการตอบข้อสอบและอีกกลุ่มหนึ่งคือ กลุ่มสนใจ (focal group) เป็นกลุ่มที่ผู้วิจัยสนใจ ศึกษาโดยทำการเปรียบเทียบกับผู้สอบกลุ่มแรก ถ้าข้อสอบทำหน้าที่ต่างกันแล้วคาดว่าผู้สอบกลุ่มนี้จะเสียประโยชน์ ในการตอบข้อสอบ ในการเปรียบเทียบผลการตอบข้อสอบจะเปรียบเทียบในทุกระดับความสามารถของผู้สอบ ย่อยของทั้งสองกลุ่มที่มีระดับความสามารถเท่ากัน ในทางปฏิบัติ มักใช้คะแนนรวมของแบบทดสอบทั้งฉบับ ซึ่งมีข้อสอบที่ศึกษารวมอยู่ด้วย เป็นเกณฑ์ในการจับคู่กลุ่มผู้สอบ ข้อสอบข้อใดที่ผู้สอบในกลุ่มที่มีความ สามารถเท่ากันทำได้ถูกต้องเท่ากัน แสดงว่า เป็นข้อสอบที่ทำหน้าที่ไม่ต่างกัน (no DIF) ระหว่างกลุ่มผู้สอบ เกณฑ์ในลักษณะดังกล่าวได้รับการวิจารณ์ว่ามีจุดอ่อน คือ การใช้คะแนนรวมของแบบทดสอบที่มีข้อสอบที่ทำ หน้าที่ต่างกันรวมอยู่ด้วย เป็นเกณฑ์ในการจับคู่ของกลุ่มผู้สอบ ก็จะรวมคะแนนของข้อสอบที่ทำหน้าที่ต่างกัน เข้าไว้ด้วย ซึ่งจะทำให้ได้เกณฑ์ที่มีความบิดเบือนไปจากความเป็นจริง ดังนั้น ผลการตรวจสอบจึงไม่ถูกต้อง ฮอลแลนด์และธาเยอร์ (Holland และ Thayer. 1988) ได้แก้ไขจุดอ่อนนี้ โดยเสนอแนะให้ใช้วิธีการวิเคราะห์ การทำหน้าที่ต่างกันของข้อสอบแบบ 2 ขั้นตอน ทั้งนี้ เพื่อให้เกณฑ์ในการจับคู่กลุ่มผู้สอบมีความบริสุทธิ์ (purification of matching criterion) ซึ่งมีขั้นตอนการวิเคราะห์ดังนี้

ขั้นตอนแรก ใช้คะแนนรวมของแบบทดสอบทั้งฉบับเป็นเกณฑ์การจับคู่กลุ่มผู้สอบย่อย 2 กลุ่ม แล้ว วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ เมื่อพบว่าข้อสอบข้อใดทำหน้าที่ต่างกันให้นำคะแนนของข้อสอบ ข้อนั้นออกจากคะแนนรวมของผู้สอบแต่ละคน

ขั้นตอนที่สอง ใช้คะแนนรวมของแบบทดสอบที่นำข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบออกไปแล้ว เป็นเกณฑ์ในการจับคู่กลุ่มผู้สอบย่อย 2 กลุ่ม จึงวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบซ้ำอีกครั้งหนึ่ง

เมื่อมีการจับคู่กลุ่มผู้สอบแล้วจะนำข้อมูลผลการตอบข้อสอบของผู้สอบย่อย 2 กลุ่มมาสร้างตาราง การณั้จร (Contingency table)

กรณีที่ข้อสอบมีการตรวจให้คะแนนแบบ 2 ค่า ตารางการณั้จรจะเป็นแบบ $2 \times 2 \times k$ โดยที่ตาราง การณั้จร 1 ตารางจะแทนคะแนนรวม 1 ระดับ ดังนั้น ถ้ามีการแบ่งคะแนนรวมเป็น k ระดับ จะต้องสร้าง ตารางการณั้จรแบบ 2×2 จำนวนทั้งหมด k ตาราง ซึ่งตารางการณั้จรแบบ 2×2 ของข้อสอบแต่ละข้อที่มี คะแนนรวมระดับ j ดังแสดงในตาราง 2

ตาราง 2 แสดงผลการตอบข้อสอบของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจที่ระดับคะแนน j

กลุ่ม	คะแนนของผลการตอบข้อสอบที่ศึกษา		
	ตอบถูก (1)	ตอบผิด (0)	รวม
อ้างอิง (R)	A_j	B_j	nR_j
สนใจ (F)	C_j	D_j	nF_j
รวม	m_{1j}	m_{0j}	T_j

เมื่อ A_j เป็น จำนวนผู้สอบกลุ่มอ้างอิงที่ตอบข้อสอบถูก ที่ระดับคะแนน j
 B_j เป็น จำนวนผู้สอบกลุ่มอ้างอิงที่ตอบข้อสอบผิด ที่ระดับคะแนน j
 C_j เป็น จำนวนผู้สอบกลุ่มสนใจที่ตอบข้อสอบถูก ที่ระดับคะแนน j
 D_j เป็น จำนวนผู้สอบกลุ่มสนใจที่ตอบข้อสอบผิด ที่ระดับคะแนน j
 m_{1j} เป็น จำนวนผู้สอบทั้งหมดที่ตอบข้อสอบถูก ที่ระดับคะแนน j
 m_{0j} เป็น จำนวนผู้สอบทั้งหมดที่ตอบข้อสอบผิด ที่ระดับคะแนน j
 nR_j เป็น จำนวนผู้สอบกลุ่มอ้างอิง ที่ระดับคะแนน j
 nF_j เป็น จำนวนผู้สอบกลุ่มสนใจ ที่ระดับคะแนน j
 T_j เป็น จำนวนผู้สอบทั้งหมด ที่ระดับคะแนน j

โดยแสดงเป็นสัดส่วนการตอบข้อสอบของผู้สอบ 2 กลุ่ม ได้ดังตาราง 3

ตาราง 3 แสดงสัดส่วนของผลการตอบข้อสอบของผู้สอบ ที่ระดับคะแนน j

กลุ่ม	คะแนนของผลการตอบข้อสอบที่ศึกษา		
	ตอบถูก(1)	ตอบผิด(0)	รวม
อ้างอิง (R)	p_{Rj}	q_{Rj}	1
สนใจ (F)	p_{Fj}	q_{Fj}	1

โดยที่ p_{Rj} คือ สัดส่วนของผู้สอบกลุ่มอ้างอิงที่ตอบข้อสอบถูก ที่ระดับคะแนน j
 q_{Rj} คือ สัดส่วนของผู้สอบกลุ่มอ้างอิงที่ตอบข้อสอบผิด ที่ระดับคะแนน j
 p_{Fj} คือ สัดส่วนของผู้สอบกลุ่มสนใจที่ตอบข้อสอบถูก ที่ระดับคะแนน j
 q_{Fj} คือ สัดส่วนของผู้สอบกลุ่มสนใจที่ตอบข้อสอบผิด ที่ระดับคะแนน j

การทดสอบสมมติฐานที่เป็นกลาง (H_0) และสมมติฐานทางเลือก (H_1) ของการทำหน้าที่ต่างกันของข้อสอบ แสดงได้ดังนี้

$$\begin{aligned}
 H_0 : p_{Rj} &= p_{Fj} && \text{สำหรับทุกระดับคะแนน } j \\
 \text{หรือ} \quad H_0 : \frac{p_{Rj}}{q_{Rj}} &= \frac{p_{Fj}}{q_{Fj}} && \text{สำหรับทุกระดับคะแนน } j \\
 H_1 : \frac{p_{Rj}}{q_{Rj}} &= \alpha \frac{p_{Fj}}{q_{Fj}} && \text{สำหรับทุกระดับคะแนน } j, \alpha \neq 1
 \end{aligned}$$

สมมติฐานที่เป็นกลาง เป็นสมมติฐานของความเป็นอิสระอย่างมีเงื่อนไขของสมาชิกของกลุ่มผู้สอบ และคะแนนของข้อสอบที่ศึกษาที่กำหนดเป็นตัวแปรการจับคู่ (matching variable)

ภายใต้สมมติฐานที่เป็นกลาง, คะแนนที่ได้ดังแสดงในตาราง 2 สรุปเป็นค่าที่คาดหวัง (expected values) ของแต่ละเซลล์ได้ ดังนี้

$$E(A_j) = n_{Rj} m_{1j} / T_j$$

$$E(B_j) = n_{Rj} m_{0j} / T_j$$

$$E(C_j) = n_{Fj} m_{1j} / T_j$$

$$E(D_j) = n_{Fj} m_{0j} / T_j$$

ค่าคาดหวังเหล่านี้เป็นความคาดหวังอย่างมีเงื่อนไข (conditional expectation) ที่กำหนดโดย n_{Rj} , n_{Fj} , m_{1j} , m_{0j}

อัตราส่วนแถมต่อรวม (Common odds ratio)

พารามิเตอร์ α เรียกว่าเป็น อัตราส่วนแถมต่อรวม ซึ่งภายใต้สมมติฐานทางเลือก ค่าของ α เป็นอัตราส่วนแถมต่อที่คำนวณได้จาก

$$\alpha = \frac{p_{Rj}}{q_{Rj}} / \frac{p_{Fj}}{q_{Fj}} = \frac{p_{Rj}q_{Fj}}{p_{Fj}q_{Rj}} \quad \text{ในทุกระดับคะแนน } j$$

ถ้า $\alpha = 1$ แสดงว่า สอดคล้องกับสมมติฐานที่เป็นกลาง นั่นคือ โอกาสของการตอบข้อสอบถูกของกลุ่มอ้างอิงและกลุ่มสนใจ มีค่าเท่ากัน แต่ถ้า $\alpha \neq 1$; กรณี $\alpha > 1$ แสดงว่า โอกาสของการตอบข้อสอบถูกของผู้สอบกลุ่มอ้างอิงมีมากกว่าผู้สอบกลุ่มสนใจ และ กรณี $\alpha < 1$ แสดงว่า โอกาสของการตอบข้อสอบถูกของผู้สอบกลุ่มสนใจมีมากกว่าผู้สอบกลุ่มอ้างอิง ซึ่ง แมนเทิลและแฮนส์เซล ได้ประมาณค่าอัตราส่วนแถมต่อรวม, α จากตาราง 2x2 ซึ่งตัวประมาณค่าคำนวณได้จาก

$$\hat{\alpha}_{MH} = \frac{\sum_{j=1}^k A_j D_j / T_j}{\sum_{j=1}^k B_j C_j / T_j} \quad (1)$$

$\hat{\alpha}_{MH}$ เป็นค่าประมาณขนาดอิทธิพลของการทำหน้าที่ต่างกันของข้อสอบ (DIF effect size) ซึ่งมีค่าอยู่ระหว่าง 0 ถึง ∞ แต่นักพัฒนาแบบทดสอบมักใช้ตัวประมาณค่าความยากของข้อสอบในการวัดที่เป็นความยากมาตรฐานของข้อสอบ (delta metric) ซึ่งมีค่าเฉลี่ยเท่ากับ 13 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 4 ในการหาค่าความยากมาตรฐานของข้อสอบ (delta : Δ) ทำได้โดยการแปลงค่าสัดส่วนของการตอบข้อสอบถูก (p) ให้เป็นคะแนน Z ด้วยการใช้อินเวอร์สของฟังก์ชันสะสมปกติ ตามวิธีการแปลงค่าเชิงเส้นตรงดังนี้

$$\Delta = 13 + 4Z$$

ดังนั้น ถ้าค่า Δ มาก แสดงว่า ข้อสอบยาก

ค่า Δ น้อย แสดงว่า ข้อสอบง่าย

ฮอลแลนด์และธาเยอร์ (Holland และ Thayer, 1985) ได้เสนอให้แปลงค่า $\hat{\alpha}_{MH}$ เป็น Δ_{MH} ซึ่งเป็นสเกลเดลต้า (delta scale : MH_{D-DIF}) ได้จากสูตร

$$\Delta_{MH} = MH_{D-DIF} = -2.35 \ln (\hat{\alpha}_{MH})$$

ซึ่งค่า MH_{D-DIF} สามารถพิจารณาเป็นค่าความยากของข้อสอบ หรือ ค่าประมาณขนาดของการทำหน้าที่ต่างกันของข้อสอบ ดังนี้

ถ้าค่า MH_{D-DIF} เป็นศูนย์ หมายความว่า ข้อสอบยากสำหรับกลุ่มอ้างอิง เท่ากับกลุ่มสนใจ แสดงว่า ข้อสอบทำหน้าที่ไม่ต่างกันสำหรับกลุ่มผู้สอบย่อยทั้ง 2 กลุ่ม

ถ้าค่า MH_{D-DIF} เป็นบวก หมายความว่า ข้อสอบยากสำหรับกลุ่มอ้างอิง มากกว่ากลุ่มสนใจ แสดงว่า ข้อสอบทำหน้าที่ต่างกัน เข้าข้างกลุ่มสนใจ

ถ้าค่า MH_{D-DIF} เป็นลบ หมายความว่า ข้อสอบยากสำหรับกลุ่มสนใจ มากกว่ากลุ่มอ้างอิง แสดงว่า ข้อสอบทำหน้าที่ต่างกัน เข้าข้างกลุ่มอ้างอิง

ค่าความคลาดเคลื่อนมาตรฐานที่ประมาณได้ของ MH_{D-DIF} สามารถคำนวณได้จากสูตร

$$SE(MH_{D-DIF}) = 2.35 \sqrt{\text{Var}[\log_e(\hat{\alpha}_{MH})]}$$

โดยที่ $\text{Var}[\log_e(\hat{\alpha}_{MH})]$ ประมาณค่าได้จาก

$$\text{Var}[\log_e(\hat{\alpha}_{MH})] = \frac{\sum_{j=1}^k U_j V_j / T_j^2}{2 \left[\sum_{j=1}^k A_j D_j / T_j \right]^2}$$

$$\text{เมื่อ } U_j = A_j D_j + \hat{\alpha}_{MH} (B_j C_j)$$

$$V_j = (A_j + D_j) + \hat{\alpha}_{MH} (B_j + C_j)$$

สถิติทดสอบที่ใช้ในการทดสอบนัยสำคัญของสมมติฐาน คือ สถิติ MH_{χ^2} ที่ระดับชั้นความเป็นอิสระเท่ากับ 1 (df = 1)

$$\text{กรณีใช้ค่า } \hat{\alpha}_{MH}; \quad H_0 : \hat{\alpha}_{MH} = 1$$

$$H_1 : \hat{\alpha}_{MH} \neq 1$$

$$\text{กรณีใช้ค่า } MH_{D-DIF}; \quad H_0 : MH_{D-DIF} = 0$$

$$H_1 : MH_{D-DIF} \neq 0$$

ซึ่งสถิติ MH_{χ^2} มีสูตรในการคำนวณดังนี้

$$MH_{\chi^2} = \frac{\left[\sum_{j=1}^k A_j - E(A_j) - .5 \right]^2}{\sum_{j=1}^k \text{var}(A_j)} \quad (2)$$

$$\text{โดยที่ } E(A_j) = n_{Rj} m_{1j} / T_j$$

$$\text{Var}(A_j) = n_{Rj} n_{Fj} m_{1j} m_{0j} / T_j^2 (T_j - 1)$$

เมื่อ $E(A_j)$ เป็น ค่าคาดหวังของจำนวนผู้สอบกลุ่มอ้างอิงที่ตอบข้อสอบถูก ที่ระดับคะแนน j

$\text{Var}(A_j)$ เป็น ค่าความแปรปรวนของจำนวนผู้สอบกลุ่มอ้างอิงที่ตอบข้อสอบถูก ที่ระดับคะแนน j

กรณีที่ข้อสอบมีการให้คะแนนแบบหลายค่า (polytomously scored item) เช่น การประเมินงานภาคปฏิบัติ (performance tasks) ได้แก่ การตัดสินคุณภาพของแฟ้มสะสมงาน (portfolio) การพิสูจน์ทางคณิตศาสตร์ (mathematical proof) การทดลองทางวิทยาศาสตร์ (scientific experiment) ข้อสอบเรียงความ (essay items) หรือข้อสอบแบบเลือกตอบที่ให้คะแนนความรู้บางส่วน (partial knowledge scoring or partial credit scoring) เป็นต้น แมนเทิลและแฮนส์เซล ได้เสนอสถิติรูปทั่วไปของ MH (Generalized MH statistic : GMH) ซึ่งเป็นวิธีการที่ขยายโดยตรงจากวิธี MH ธรรมดา ในกรณีที่ลำดับชั้นการตอบ(response categories) มีมากกว่า 2 ลำดับชั้น ($T > 2$) และลำดับชั้นการตอบไม่จำเป็นต้องมีการจัดเรียง ข้อมูลการตอบข้อสอบจะถูกจัดลงตารางการณัจร $2 \times T \times k$ เมื่อ T เป็นจำนวน ลำดับชั้นการตอบและ k เป็นจำนวนของระดับของการแบ่งชั้นตัวแปรการจับคู่ ในแต่ละระดับทั้งหมด k ระดับ ดังตาราง 4

ตาราง 4 แสดงผลการตอบข้อสอบของกลุ่มผู้สอบ ในแต่ละลำดับชั้นคะแนน

กลุ่ม	คะแนนของผลการตอบข้อสอบ					รวม
	Y_1	Y_2	Y_3	...	Y_T	
อ้างอิง (R)	n_{R1k}	n_{R2k}	n_{R3k}	...	n_{RTk}	n_{R+k}
สนใจ (F)	n_{F1k}	n_{F2k}	n_{F3k}	...	n_{FTk}	n_{F+k}
รวม	n_{+1k}	n_{+2k}	n_{+3k}	...	n_{+Tk}	n_{++k}

ค่า $y_1, y_2, y_3, \dots, y_T$ แทน คะแนนในลำดับการตอบที่ได้จากข้อสอบข้อนั้น
(เป็นจำนวนเต็ม 1, 2, ..., T)
 n_{RTk} และ n_{FTk} แทน จำนวนของสมาชิกในกลุ่ม R และ F ตามลำดับ
ณ ระดับตัวแปรจับคู่ที่ k และได้คะแนน y_T
เครื่องหมาย + แทน ผลรวมตลอดจนถึงที่บ่งชี้ไว้ เช่น n_{F+k} แทน จำนวนสมาชิกของกลุ่มสนใจทั้งหมด ในระดับชั้นที่ k ซึ่งเวกเตอร์ของความถี่ของการตอบของกลุ่ม g มีการแจกแจงแบบ multivariate hypergeometric ด้วยพารามิเตอร์ n_{F+k} นั่นคือ
$$n_{Fk} = (n_{F1k}, n_{F2k}, \dots, n_{FTk})$$

และ $n_{+k} = (n_{+1k}, n_{+2k}, \dots, n_{+Tk})$
สมมติฐาน ; $H_0 : \alpha_k = \alpha = 1 ; k = 1, 2, \dots, K$
 $H_1 : \alpha_k = \alpha \neq 1 ; k = 1, 2, \dots, K$
ซึ่งแมนเทิลและแฮนส์เซล ได้กำหนด $\hat{\alpha}_{GMH}$ เป็นตัวประมาณค่าอัตราส่วนแถมต่อรวม, α
ดังนั้น การทดสอบสมมติฐาน; $H_0 : \hat{\alpha}_{GMH} = 1$
 $H_1 : \hat{\alpha}_{GMH} \neq 1$
สถิติทดสอบที่ใช้ในการทดสอบนัยสำคัญของสมมติฐาน คือ สถิติ GMH_{χ^2} ซึ่งมีการแจกแจงแบบไค-สแควร์ ที่มี $df = T-1$ ค่าพจน์ได้จากสูตร

$$GMH_{\chi^2}^2 = [\sum A_k - \sum E(A_k)]' [\sum V(A_k)]^{-1} [\sum A_k - \sum E(A_k)]$$

(3)

โดยที่ $A'_k = (n_{R1k}, n_{R2k}, \dots, n_{R(T-1)k})$

$$E(A'_k) = n_{R+k} n'_k / n_{++k}$$

$$n'_k = (n_{+1k}, n_{+2k}, \dots, n_{+(T-1)k})$$

$$V(A_k) = n_{R+k} n_{F+k} \left(\frac{n_{++k} \text{diag}(n_k) - n_k n'_k}{n_{++k}^2 (n_{++k} - 1)} \right)$$

และ $\text{diag}(n_k)$ เป็นเมตริกซ์แนวทแยง (diagonal matrix) มิติ $(T-1) \times (T-1)$ ที่มีสมาชิกเป็น n_k

เมื่อ A_k และ $E(A_k)$ เป็นเวกเตอร์ที่ยาว $T-1$ ที่สมนัยกับ $T-1$ ไต ๆ ของลำดับชั้นการตอบ T

และ $V(A_k)$ เป็นเมตริกซ์ความแปรปรวน-ความแปรปรวนร่วม (variance-covariance matrix) มิติ $(T-1) \times (T-1)$

3. วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี **Polytomous SIBTEST**

เชลลีและสตาต์ (Shealy & Stout, 1993 : 159-194) ได้เสนอวิธี SIBTEST (Simultaneous Items Bias Test) เพื่อใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (Differential item functioning : DIF) การทำหน้าที่ต่างกันของแบบทดสอบ (Differential test functioning : DTF) และการทำหน้าที่ต่างกันของกลุ่มข้อสอบ (Differential bundle functioning : DBF) เป็นวิธีการที่พัฒนามาจากแนวคิดของการตรวจสอบการทำหน้าที่ต่างกันของแบบทดสอบบนพื้นฐานของทฤษฎีการตอบข้อสอบแบบหลายมิติ (multidimensional IRT modeling of DTF) มีรูปแบบเป็นนัยพารามेटริกแบบตัวแปรแฝง (latent - variable non-parametric) ไม่ต้องใช้ฟังก์ชันการตอบข้อสอบ หรือ การประมาณค่าความสามารถแฝง สามารถวิเคราะห์ได้ทั้งในแบบทดสอบที่เป็นมิติเดียว (unidimensional test) และแบบทดสอบที่เป็นหลายมิติ (multidimensional test) (Stout, Li and Nandakumar, 1997 : 195) จุดเด่นของวิธีการ SIBTEST คือ มีการคำนวณที่ง่าย ไม่ซับซ้อน เสียค่าใช้จ่ายไม่มาก ไม่จำเป็นต้องใช้กลุ่มตัวอย่างขนาดใหญ่ ใช้ได้ดีสำหรับการตรวจสอบการทำหน้าที่ต่างกันแบบสม่ำเสมอ (uniform DIF) และมีทิศทางเดียว (unidirectional DIF) นอกจากนั้น ยังสามารถใช้ได้กับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบ 2 ค่า (dichotomous DIF) และที่มีการให้คะแนนแบบหลายค่า (polytomous DIF) (Chang, Mazzeo and Roussos, 1995 cited in Potenza and Dorans, 1995 : 31)

ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี SIBTEST ในแบบทดสอบที่เป็นมิติเดียวจะต้องมีข้อตกลงว่ามีมิติการวัด 2 มิติ $\{\theta = (\theta, \eta)\}$ คือ มิติหนึ่ง เป็นความสามารถเป้าหมายหรือคุณลักษณะแฝงเพียงลักษณะเดียวที่ต้องการวัด (target ability; θ) และ อีกมิติหนึ่ง เป็นคุณลักษณะแฝงแทรกซ้อนหรือความสามารถแทรกซ้อนที่ไม่ต้องการวัด (nuisance ability; η) ตัวอย่างเช่น แบบทดสอบวัดความสามารถในเรื่องคำศัพท์ภาษาอังกฤษ ข้อสอบบางข้ออาจถามความรู้พิเศษสำหรับเพศชาย เช่น ความรู้เรื่องกีฬา ส่วนข้อสอบบางข้ออาจถามความรู้พิเศษสำหรับเพศหญิง เช่น ความรู้ในเรื่องงานบ้าน ในสถานการณ์เช่นนี้ทักษะในเรื่องคำศัพท์ภาษาอังกฤษ เป็นความสามารถเป้าหมายที่ต้องการวัด ซึ่งแทนด้วย θ ส่วนความรู้เรื่องกีฬา และความรู้เรื่องงานบ้าน เป็นความสามารถแทรกซ้อนที่ไม่ต้องการวัด ซึ่งแทนด้วย η_1 และ η_2 ตามลำดับ ข้อสอบทุกข้อในแบบทดสอบ ถือว่าวัดความสามารถเป้าหมาย ส่วนข้อสอบที่ทำหน้าที่ต่างกันจะวัดทั้งความสามารถเป้าหมายและความสามารถแทรกซ้อน 1 หรือ มากกว่า 1 ความสามารถ

ให้ U แทน ผลการตอบข้อสอบได้มาจากการสุ่มจากกลุ่มผู้ตอบในแต่ละกลุ่ม
 ซึ่ง $U = (U_1, U_2, \dots, U_N)$ โดยที่ $U_i = 1$ ถ้าตอบข้อสอบข้อที่ i ถูกต้อง
 และ $U_i = 0$ ถ้าตอบข้อสอบข้อที่ i ผิด

N เป็นจำนวนข้อสอบ

$P_i(\theta)$ แทน IRF ของข้อสอบข้อที่ i ที่ขึ้นอยู่กับความสามารถ θ เพียงอย่างเดียว

$P_i(\theta, \eta)$ แทน IRF ของข้อสอบข้อที่ i ที่ขึ้นอยู่กับความสามารถ θ และ η
 แสดงได้ดังนี้

$$P_i(\theta) = C_i + \frac{(1 - C_i)}{1 + \exp[-1.7 a_{i\theta}(\theta - b_{i\theta})]}, \quad i = 1, \dots, N$$

$$P_i(\theta, \eta) = C_i + \frac{(1 - C_i)}{1 + \exp\{-1.7[a_{i\theta}(\theta - b_{i\theta}) + a_{i\eta}(\eta - b_{i\eta})]\}}, \quad i = n+1, \dots, N$$

ซึ่ง IRF ของทั้งสองลักษณะ ถือว่าเป็นไปตามข้อตกลงเบื้องต้นตามทฤษฎี IRT ในเรื่องความเป็นอิสระต่อการตอบข้อสอบ (local independence) ของ U มีลักษณะเป็น monotonicity และมีคุณสมบัติไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบ (invariance) เมื่อให้ IRF แทน ด้วย $P_i(\theta, \eta)$ สามารถกำหนดในรูปของความน่าจะเป็น ได้ดังนี้ (Li and Stout, 1996 : 650)

$$P[U | (\Theta = \theta, \eta = \eta)] = \prod_{i=1}^N P_i(\theta, \eta)^{U_i} [1 - P_i(\theta, \eta)]^{1 - U_i}$$

หลักในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี SIBTEST นั้นจะเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มอ้างอิง (R) และกลุ่มสนใจ (F) โดยแบ่งแบบทดสอบออกเป็น 2 ชุดย่อย (subtests) คือ (1) แบบทดสอบย่อยที่มีความเที่ยงตรง (valid subtest) หรือเป็นแบบทดสอบที่ใช้ในการจับคู่เปรียบเทียบ (matching subtest) ซึ่งแบบทดสอบย่อยชุดนี้ประกอบด้วยข้อสอบที่ตั้งใจวัดความสามารถเป้าหมายเพียงอย่างเดียว (2) แบบทดสอบย่อยที่ต้องการศึกษา (studied subtest) ซึ่งประกอบด้วยข้อสอบที่วัดความสามารถที่ต้องการวัดและความสามารถที่ไม่ต้องการวัด ถ้าแบบทดสอบย่อยชุดแรกมีจำนวน n ข้อแล้วแบบทดสอบย่อยชุดที่ 2 จะมีจำนวน $N-n$ ข้อ เมื่อ N เป็นจำนวนข้อสอบทั้งหมด

กรณีที่ เป็นแบบทดสอบที่มีการให้คะแนนแบบ 2 ค่า

$$\text{ให้} \quad X = \sum_{i=1}^n U_i$$

$$Y = \sum_{i=n+1}^N U_i$$

เมื่อ X เป็นคะแนนรวมจากแบบทดสอบที่มีความเที่ยงตรง

Y เป็นคะแนนรวมจากแบบทดสอบที่ต้องการศึกษา

U_i เป็นผลการตอบข้อสอบข้อที่ i (ตอบถูกได้ 1 ตอบผิดได้ 0)

นำคะแนนเฉลี่ยจากผลการตอบข้อสอบในแบบทดสอบย่อยที่ต้องการศึกษาระหว่างผู้สอบกลุ่มอ้างอิง และกลุ่มสนใจ ที่มีความสามารถระดับเดียวกันมาจับคู่เปรียบเทียบกัน โดยพิจารณาจากคะแนนรวมของแบบทดสอบย่อยที่เที่ยงตรงที่เท่ากัน ($x = k$) ได้ดังนี้

$$\bar{Y}_{Rk} - \bar{Y}_{Fk} \quad ; k = 0, 1, \dots, n$$

เมื่อ $\bar{Y}_{Rk}$ เป็นค่าเฉลี่ยของคะแนนรวมจากแบบทดสอบย่อยที่ต้องการศึกษา

ของผู้สอบกลุ่มอ้างอิง ซึ่งได้คะแนน $x = k$

$\bar{Y}_{Fk}$ เป็นค่าเฉลี่ยของคะแนนรวมจากแบบทดสอบย่อยที่ต้องการศึกษา

ของผู้สอบกลุ่มสนใจ ซึ่งได้คะแนน $x = k$

K เป็นคะแนนรวมจากแบบทดสอบย่อยที่มีความเที่ยงตรง

ค่า $\bar{Y}_{Rk} - \bar{Y}_{Fk}$ เป็นความแตกต่างของผลการตอบข้อสอบจากแบบทดสอบย่อยที่ต้องการศึกษา ระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจ ที่มีระดับความสามารถที่ต้องการวัด (θ) เท่ากัน

ในกรณีที่ข้อสอบที่ต้องการศึกษาทำหน้าที่ไม่ต่างกัน ค่า $\bar{Y}_{Rk} - \bar{Y}_{Fk} = 0$

แต่ถ้า $\bar{Y}_{Rk} - \bar{Y}_{Fk} \neq 0$ แสดงว่า ข้อสอบที่ศึกษาทำหน้าที่ต่างกัน

สมมติฐานในการทดสอบ

$$H_0 : \beta_U = 0$$

$$H_1 : \beta_U > 0$$

ซึ่ง β_U เป็นดัชนีบ่งชี้การทำหน้าที่ต่างกันของข้อสอบ ประมาณค่าได้จากสูตร

$$\hat{\beta}_U = \sum_{k=0}^n \hat{p}_k (\bar{Y}_{Rk} - \bar{Y}_{Fk}) \quad (4)$$

$\hat{\beta}_U$ เป็นค่าสถิติที่ใช้ประมาณค่าปริมาณของการทำหน้าที่ต่างกันของข้อสอบแบบทิศทางเดียว

ของ β_U

เมื่อ $\hat{p}_k$ เป็นสัดส่วนของผู้สอบกลุ่มสนใจที่ได้คะแนนจากแบบทดสอบที่มีความเที่ยงตรง $x = k$

สถิติทดสอบที่ใช้ทดสอบสมมติฐานที่เป็นกลาง (no DIF)

$$B = \frac{\hat{\beta}_U}{\hat{\sigma}(\hat{\beta}_U)} \quad (5)$$

$$\text{โดยที่ } \hat{\sigma}(\hat{\beta}_U) = \left[\sum_{k=0}^n \hat{p}_k^2 \left[\frac{1}{J_{Rk}} \hat{\sigma}^2(y|k,R) + \frac{1}{J_{Fk}} \hat{\sigma}^2(y|k,F) \right] \right]^{1/2}$$

เมื่อ $\hat{\sigma}(\hat{\beta}_U)$ เป็นความคลาดเคลื่อนมาตรฐานในการประมาณค่าของ $\hat{\beta}_U$

$\hat{\sigma}^2(y|k,g)$ เป็นค่าประมาณความแปรปรวนของคะแนนจากแบบทดสอบที่ศึกษาสำหรับกลุ่มผู้สอบ g (R หรือ F) ที่ได้คะแนนรวม $x = k$ จากแบบทดสอบที่มีความเที่ยงตรง

และ J_{Rk} และ J_{Fk} เป็นจำนวนผู้สอบของกลุ่มอ้างอิงและกลุ่มสนใจ ตามลำดับ ที่ได้คะแนนรวม $x = k$ จากแบบทดสอบที่มีความเที่ยงตรง

สถิติที่ใช้ในการทดสอบ B มีการแจกแจงปกติมาตรฐาน $N(0,1)$ เมื่อ $B = 0$ และจะปฏิเสธ H_0 ถ้าผลการทดสอบปรากฏว่า $B > Z_\alpha$ อย่างมีนัยสำคัญที่ระดับ α โดยที่ $\alpha = P[N(0,1) > Z_\alpha]$ นั่นคือข้อสอบทำหน้าที่ต่างกัน โดยจะเข้าข้างผู้สอบกลุ่มสนใจเมื่อ B มีค่าเป็นบวก และจะเข้าข้างผู้สอบกลุ่มอ้างอิงเมื่อ B มีค่าเป็นลบ

วิธีการ SIBTEST ดังกล่าวข้างต้นเป็นวิธีการที่ใช้กับข้อสอบที่ให้คะแนนแบบ 2 ค่า (dichotomously scored items) เชลลีและสตาต์ (Shealy and Stout, 1993) ได้พัฒนาขยายวิธีการ SIBTEST ให้ใช้ได้กับข้อสอบที่มีการให้คะแนนแบบหลายค่า (polytomously scored items) มีรายละเอียดดังนี้

ให้ y เป็นคะแนนจากข้อสอบที่ศึกษา ซึ่งมีลำดับชั้นที่จัดเรียงแล้ว (ordered categories) $m+1$ ลำดับชั้น ($y = k ; 0 \leq k \leq m$)

ให้ $P_{kg}(\theta)$ แทน ฟังก์ชันการตอบลำดับชั้นข้อสอบ (item category response function : ICRF) ซึ่งเป็นความน่าจะเป็นของการได้คะแนน k ของผู้สอบที่มีความสามารถ θ จากกลุ่ม g ($g = R$ หรือ F)

การถดถอยของคะแนนสอบบนความสามารถ สามารถนิยามเป็น ผลรวม ICRF_g ที่ให้ค่านำหนัก (weighted sum of ICRF_g) ได้ดังนี้

$$E_g[y|\theta] = \sum_{k=1}^m k P_{kg}(\theta)$$

ซึ่งการถดถอยนี้เป็น ฟังก์ชันการตอบข้อสอบ (item response function:IRF) ในกรณีที่ข้อสอบมีการให้คะแนนแบบ 2 ค่า สามารถพิจารณาเป็นข้อสอบที่มีการให้คะแนนแบบหลายค่า ที่มี $m = 1$

สำหรับโมเดลตามทฤษฎี IRT ที่เป็นแบบ 2 ค่า นั้น โครงสร้างของข้อสอบข้อหนึ่งสามารถหา IRF ตามที่กำหนดได้อย่างชัดเจน ส่วนโมเดลตามทฤษฎี IRT ที่เป็นแบบหลายค่า นั้น โครงสร้างของข้อสอบที่มี $m+1$ ลำดับชั้น สามารถหาฟังก์ชันการตอบลำดับชั้นข้อสอบ (ICRF) ได้ $m(\text{ICRF})$

เมื่อ $P_{kR}(\theta)$ และ $P_{kF}(\theta)$ เป็นฟังก์ชันการตอบลำดับชั้นข้อสอบ (ICRF) ของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจ ตามลำดับ นิยามของ null-DIF แบบตัวแปรแฝง เป็นดังนี้

$$P_{kR}(\theta) = P_{kF}(\theta) ; k = 1, \dots, m \quad (6)$$

สมการ 6 ใช้ได้กับโมเดลการตอบข้อสอบที่เป็นแบบหลายค่า ซึ่งโมเดลที่นิยมใช้มากที่สุด 3 โมเดล คือ พาเชียล เครดิต โมเดล (PCM), พาเชียล เครดิต โมเดล รูปทั่วไป (GPCM) และ เกรด เรสพอนส์ โมเดล (GRM) การพิจารณาการทำหน้าที่ต่างกันของข้อสอบโดยพิจารณาเพียงฟังก์ชันการตอบ (IRF) ที่ต่างกันระหว่างความสามารถของกลุ่มเป็นการเพียงพอ เนื่องจากการเปรียบเทียบเพียงฟังก์ชันการตอบ (IRF) ไม่มีสารสนเทศที่สูญหาย

ให้ $x_1, x_2, \dots, x_n$ เป็นคะแนนจากข้อสอบที่ใช้ในการจับคู่ n ข้อ

$m_1, m_2, \dots, m_n$ เป็นคะแนนที่เป็นไปได้สูงสุด (maximum possible scores)

สำหรับ $x_1, x_2, \dots, x_n$ ตามลำดับ

ถ้า $m_i = 1 ; x_i$ คือข้อสอบที่ให้คะแนนเป็น 2 ค่า (dichotomous)

$$X = \sum_{j=1}^n x_j \quad \text{เป็นคะแนนรวมที่ใช้ในการจับคู่ ; } x = 0, 1, \dots, n_H$$

$$\text{เมื่อ } n_H = \sum_{j=1}^n m_j \quad \text{เป็นคะแนนของการจับคู่ที่เป็นไปได้สูงสุด}$$

$$\bar{y}_{gk} \quad \text{เป็นคะแนนเฉลี่ยจากข้อสอบที่ใช้ศึกษาสำหรับผู้สอบในกลุ่ม } g \text{ ทั้งหมด}$$

$$(g = R \text{ หรือ } F) \text{ ที่ได้ } x = k$$

และ P_{Fk} เป็นสัดส่วนของผู้สอบในกลุ่มสนใจที่ได้คะแนน $x = k$ จากการสอบเพื่อจับคู่ $x_1, x_2, \dots, x_n$ นั่นคือ

$$P_{Fk} = \frac{N_{Fk}}{N_F}$$

$$\text{เมื่อ } N_F \quad \text{เป็นจำนวนของผู้สอบทั้งหมดในกลุ่มสนใจ}$$

$$\text{และ } N_{Fk} \quad \text{เป็นจำนวนของผู้สอบทั้งหมดในกลุ่มสนใจที่ได้คะแนนการจับคู่ } x = k$$

$$P_k \quad \text{เป็นสัดส่วนของผู้สอบทั้งหมดที่ได้คะแนน } x = k \text{ จากการสอบเพื่อจับคู่}$$

นั่นคือ

$$P_k = \frac{N_k}{N}$$

$$\text{เมื่อ } N \quad \text{เป็นจำนวนผู้สอบทั้งหมด}$$

$$\text{และ } N_k = N_{Rk} + N_{Fk}$$

เนื่องจากวิธีการ SIBTEST มีพื้นฐานแนวคิดของทฤษฎีการตอบข้อสอบ (IRT) การนิยามการทำหน้าที่ต่างกันของข้อสอบแบบตัวแปรแฝง (Latent-variable DIF) ก็คือ

$$E_R[Y|\theta] \neq E_F[Y|\theta]$$

และปริมาณการทำหน้าที่ต่างกันของข้อสอบ ณ ความสามารถ θ หาได้โดย

$$B_o(\theta) = E_R[Y|\theta] - E_F[Y|\theta]$$

ส่วนดัชนีของการทำหน้าที่ต่างกัน (Index of DIF) คือ

$$\beta = \int B_o(\theta) f_F(\theta) d\theta$$

แต่ถ้าพิจารณาตามทฤษฎีการทดสอบแบบมาตรฐานเดิม (CTT : $X = T + E$) แล้ว T เป็นคะแนนจริงที่ใช้ในการจับคู่ ส่วน $E = X - T$ นั้น เป็นความคลาดเคลื่อนของการวัดและถือว่ามีความเป็นศูนย์ทั้ง 2 กลุ่ม ให้ $f_g(t)$ เป็นความหนาแน่นของคะแนนจริงที่ใช้ในการจับคู่ในกลุ่ม g แล้ว

การถดถอยของคะแนนของข้อสอบที่ศึกษาบนคะแนนสอบของข้อสอบที่ใช้ในการจับคู่ของกลุ่ม g แทนด้วย

$$E_g[Y|t] = E[Y|T = t, G = g]$$

กรณีที่ไม่มีการทำหน้าที่ต่างกันของข้อสอบ (no DIF) แล้ว $E_R[Y|t] = E_F[Y|t]$ นั่นคือ ปริมาณการทำหน้าที่ต่างกันของข้อสอบ ณ ความสามารถ t หาได้โดย

$$B(t) = E_R[Y|t] - E_F[Y|t]$$

ส่วนดัชนีของการทำหน้าที่ต่างกัน (Index of DIF) คือ

$$\beta = \int B(t) f_F(t) dt$$

สถิติที่ใช้ประมาณค่าปริมาณของการทำหน้าที่ต่างกันของข้อสอบของ β คือ

$$\hat{\beta} = \sum_{k=0}^{n_H} P_k d_k \quad (7)$$

เมื่อ $d_k = \bar{y}_{Rk} - \bar{y}_{Fk}$; $k = 0, \dots, n_H$

ซึ่ง เชลลีและสตาต์ (Shealy and Stout. 1993) ได้ศึกษาจากข้อมูลจำลองโดยใช้ P_k แทน P_{Fk} เป็นการถ่วงน้ำหนัก พบว่า สามารถควบคุม ความผิดพลาดประเภท 1 ได้ดีกว่า

สถิติทดสอบ

$$B = \frac{\hat{\beta}}{\hat{\sigma}(\hat{\beta})} \quad (8)$$

$$\text{เมื่อ } \hat{\sigma}(\hat{\beta}) = \left[\sum_{k=0}^{n_H} P_k^2 \left(\frac{\hat{\sigma}^2(y|k,R)}{N_{Rk}} + \frac{\hat{\sigma}^2(y|k,F)}{N_{Fk}} \right) \right]^{1/2}$$

และ $\hat{\sigma}^2(y|k,g)$ เป็นความแปรปรวนของคะแนนจากข้อสอบที่ใช้ศึกษาของผู้สอบในกลุ่ม g ($g = R$ หรือ F) ที่มีคะแนนรวมจากข้อสอบที่ใช้จับคู่ $x = k$

สถิติที่ใช้ในการอ้างอิงการทำหน้าที่ต่างกันของข้อสอบ มักมีปัญหาในกรณีที่มีความแตกต่างของการแจกแจงความสามารถที่ต้องการวัดระหว่างกลุ่มผู้สอบ ที่เรียกว่า ผลกระทบ (Impact) คือ ผู้สอบกลุ่มอ้างอิงมีความสามารถที่ต้องการวัดสูงกว่าผู้สอบกลุ่มสนใจ ก่อให้เกิดอัตราความผิดพลาดประเภทที่ 1 ที่เฟ้อหรือสูงเกินไป (inflate) ซึ่งในความเป็นจริงข้อสอบทำหน้าที่ไม่ต่างกัน ดังนั้น จึงมีความจำเป็นต้องแก้ไขความแตกต่างของการแจกแจงความสามารถที่ต้องการวัดด้วยวิธีการปรับแก้ค่าการถดถอย (regression correction) (Shealy & Stout. 1993 :170) เพื่อแก้ไขและกำจัดอิทธิพลการเกิดความผิดพลาดประเภท 1 ที่เกิดขึ้นมาากันเนื่องมาจากผลกระทบ โดยการแปลงค่า $\bar{y}_{Rk}$ และ $\bar{y}_{Fk}$ ให้เป็น $\bar{y}_{Rk}^*$ และ $\bar{y}_{Fk}^*$ ตามลำดับที่ละคู่ ดังนี้

$$\bar{y}_{gk}^* = \bar{y}_{gk} + \hat{M}_{gk} (\hat{V}(k) - \hat{V}_g(k)) \quad (9)$$

โดยที่

$$\hat{M}_{gk} \equiv \frac{\bar{y}_{g,k+1} - \bar{y}_{g,k-1}}{\hat{V}_g(k+1) - \hat{V}_g(k-1)}$$

$$\hat{V}(k) \equiv \frac{1}{2}(\hat{V}_R(k) + \hat{V}_F(k))$$

เมื่อ $\bar{y}_{gk}^*$ แทน ค่าเฉลี่ยที่ปรับแก้แล้วของคะแนนรวมจากแบบทดสอบที่ต้องการศึกษา
ในกลุ่มผู้สอบ g

$\hat{M}_{gk}$ แทน ค่าประมาณคร่าว ๆ ของค่าเบี่ยงเบนของคะแนนจริงจากแบบทดสอบ
ที่ต้องการศึกษาที่มีต่อฟังก์ชันของคะแนนจริงจากแบบทดสอบที่มีความเที่ยงตรง

$\hat{V}(k)$ แทน ค่าประมาณของคะแนนจากแบบทดสอบที่มีความเที่ยงตรง ซึ่งได้
คะแนนสังเกต $x = k$

$\hat{V}_g(k)$ แทน ค่าประมาณการถดถอยของคะแนนจริงจากแบบทดสอบที่มีความเที่ยงตรง
ของผู้สอบกลุ่ม g ซึ่งได้คะแนนสังเกต $x = k$

และ adjusted d_k^* ได้มาโดยการแปลงค่าเชิงเส้น (linear transformation) ดังนี้

$$d_k^* = \bar{y}_{Rk}^* - \bar{y}_{Fk}^* ; k = 0, 1, \dots, n_H \quad (10)$$

ดังนั้น สถิติที่ใช้ประมาณค่าปริมาณการหาหน้าที่ต่างกันของข้อสอบของ β จึงเปลี่ยนเป็น

$$\hat{\beta}^* = \sum_{k=0}^{n_H} p_k d_k^* \quad (11)$$

H_0 เป็นจริง เมื่อ B มีการแจกแจงเป็นปกติมาตรฐาน $N(0,1)$ และจะปฏิเสธ H_0 เมื่อ $|B| > 1.96$
ที่ระดับนัยสำคัญ .05

4. วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี GPCM (Generalized Partial Credit Model : GPCM)

ทฤษฎีการตอบข้อสอบ (Item Response Theory : IRT) หรือทฤษฎีคุณลักษณะแฝง (Latent Trait Theory) เป็นทฤษฎีที่ได้รับการพัฒนามานาน เริ่มตั้งแต่ปี 1916 โดยบิเนย์และไซมอน (Binet & Simon) ได้เริ่มสร้างกราฟแสดงความสัมพันธ์ระหว่างผลการสอบกับตัวแปรอิสระอื่น ๆ และใช้กราฟนั้นเป็นเครื่องมือในการพัฒนาแบบทดสอบ ต่อมาปี 1936 ริชาร์ดสัน (Richardson) ได้แสดงความสัมพันธ์ระหว่างพารามิเตอร์ของโมเดลการตอบข้อสอบ กับ พารามิเตอร์ของทฤษฎีการทดสอบมาตรฐานเดิม ซึ่งนับเป็นวิธีแรกในการประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบข้อสอบ ในช่วงปี 1943-1944 โลว์ลีย์ (Lawley) ได้เสนอวิธีการใหม่ในการประมาณค่าพารามิเตอร์ ซึ่งมีหลักการว่า ผลการสอบของผู้สอบจากแบบทดสอบใด ๆ ขึ้นอยู่กับความสามารถของผู้สอบ (Ability or Skill) ในปี 1952 ลอร์ด (Lord) ได้เสนอทฤษฎีในรูปโค้งลักษณะข้อสอบ (Item Characteristic Curve : ICC) ซึ่งมีลักษณะเป็นโค้งปกติสะสม เรียกว่า Normal Ogive Model ซึ่งมีพารามิเตอร์ 2 ตัว คือ ค่าความยาก และค่าอำนาจจำแนก ในทางปฏิบัติ โมเดลนี้มีการคำนวณที่ยุ่งยากและขาดโปรแกรมคอมพิวเตอร์ที่มีประสิทธิภาพในการวิเคราะห์ข้อมูล ทำให้การทำงานของ ลอร์ด ชะงักไประยะหนึ่งและหันมาสนใจอีกครั้ง ในปี 1965 ต่อมาปี 1960 ราสช์ (Rasch) ศึกษาทฤษฎีนี้เช่นกันได้เสนอแนวคิดในรูปของพารามิเตอร์ตัวเดียว คือค่าความยากของข้อสอบเท่านั้น ซึ่งเรียกว่า Rasch Model ในปี 1968 เบอร์นบอม (Birnbaum) ได้เสนอให้ใช้ Logistic Model ที่ใช้พารามิเตอร์ 2 ตัว คือ ค่าความยากและค่าอำนาจจำแนก ซึ่งเป็นโมเดลที่คำนวณง่ายกว่าของลอร์ด และยังใช้ได้กับพารามิเตอร์ตัวเดียวและพารามิเตอร์ 3 ตัว ทำให้โมเดลนี้มีผู้นิยมนำไปใช้ในการศึกษาอย่างแพร่หลาย และ ได้รับการพัฒนาขึ้นเรื่อยมา (Hambleton & Swaminathan, 1985 : 4-8)

ทฤษฎีการตอบข้อสอบ เป็นทฤษฎีการวัดที่อธิบายถึงความสัมพันธ์ระหว่างคุณลักษณะภายในหรือความสามารถที่มีอยู่ในตัวบุคคลกับพฤติกรรมการตอบข้อสอบของบุคคลนั้น ในรูปฟังก์ชันทางคณิตศาสตร์ ทฤษฎีนี้มีพื้นฐานความเชื่อว่า พฤติกรรมการตอบข้อสอบของผู้สอบ ถูกกำหนดโดยคุณลักษณะภายในหรือความสามารถที่อยู่ภายในตัวบุคคล ซึ่งไม่สามารถสังเกตได้โดยตรง (Lord and Novick, 1968 : 358) เขียนในรูปฟังก์ชันทางคณิตศาสตร์ ได้ดังนี้

$$Y = f(\theta)$$

เมื่อ y แทน คะแนนที่ได้จากการตอบข้อสอบ ซึ่งเป็นพฤติกรรมการตอบข้อสอบ

θ แทน คุณลักษณะภายในหรือความสามารถที่อยู่ภายในตัวบุคคล

ฟังก์ชันนี้เรียกว่า ฟังก์ชันการตอบข้อสอบ (Item Response Function : IRF) หรือ ฟังก์ชันคุณลักษณะข้อสอบ (Item Characteristic Function) ซึ่งฟังก์ชันนี้ มีลักษณะเป็นโค้งสะสมเพิ่มขึ้นทางเดียว (monotonically increasing function) ที่แสดงว่า ผู้สอบที่มีคะแนนการวัดคุณลักษณะแฝงที่สูงกว่า จะมีความน่าจะเป็นที่คาดหวังในการตอบข้อสอบได้ถูกต้อง สูงกว่า ผู้สอบที่ได้คะแนนการวัดคุณลักษณะแฝงนั้นต่ำกว่า (Hambleton and Swaminathan, 1985 : 13)

ทฤษฎีการตอบข้อสอบ มีข้อตกลงเบื้องต้นที่สำคัญ ดังนี้ (Lord, 1980 : 19-21; Hambleton and Swaminathan, 1985 : 16-31)

1. ความเป็นมิติเดียว (Unidimensionality) หมายความว่า ถ้าความสามารถหรือคุณลักษณะของคนที่อยู่หลายอย่าง ซึ่งความสามารถหรือคุณลักษณะแต่ละอย่างนี้ ต่างส่งผลต่อการตอบข้อสอบต่าง ๆ ที่

รวมกันเป็นแบบทดสอบ ถ้าผลการตอบข้อสอบหรือคะแนนสอบนั้น สามารถอธิบายได้เพียงความสามารถเดียว (only one ability) ก็เพียงพอแล้ว ข้อตกลงนี้ เรียกว่า มีความเป็นมิติเดียว (Hambleton and Swaminathan, 1985 :16) โดยทั่วไปแล้ว เป็นการยากที่จะสร้างแบบทดสอบให้เป็นไปตามข้อตกลงเบื้องต้นข้อนี้ แอมเบิลตันและสวามินาธาน (Hambleton & Swaminathan, 1985) ได้กล่าวว่า ข้อตกลงเบื้องต้นข้อนี้ไม่ได้เข้มงวดมากนัก ลักษณะที่เป็นไปได้ คือ พยายามเลี่ยงไปใช้ข้อตกลงที่ผ่อนคลายเป็นพิเศษ โดยพิจารณามิติเด่น (dominant component) เพียงมิติเดียว ก็น่าจะเพียงพอแล้ว นั่นคือ สร้างให้แบบทดสอบสามารถวัดความสามารถ หรือ คุณลักษณะที่ต้องการวัดเป็นส่วนใหญ่ ถึงแม้ว่า อาจมีคุณลักษณะอื่นเข้ามาเกี่ยวข้องบ้างก็เป็นสิ่งที่หลีกเลี่ยงไม่ได้ ความเป็นมิติเดียวในลักษณะนี้ เรียกว่า มิติหลักที่สำคัญ (essential unidimension) แต่ถ้าผลการตอบข้อสอบหรือคะแนนสอบนั้น ต้องอธิบายด้วยความสามารถหลายอย่าง เรียกข้อตกลงนี้ว่า ความเป็นหลายมิติ (Multidimensionality)

วิธีการตรวจสอบความเป็นมิติเดียวของแบบทดสอบทำได้หลายวิธี ซึ่งวิธีที่นิยมใช้ คือ วิธีการวิเคราะห์องค์ประกอบ (factor analysis) โดยวิเคราะห์องค์ประกอบที่สำคัญ เพื่อพิจารณาค่าความแปรปรวนสูงสุดขององค์ประกอบหลักตัวแรก ว่ามากกว่าองค์ประกอบอื่น หรือ ค่าไอเกน (Eigen Value) ค่าสูงสุดว่าแตกต่างจากค่าไอเกนขององค์ประกอบอื่นๆ อย่างชัดเจนหรือไม่ ถ้ามีมากกว่าหรือต่างอย่างชัดเจน แสดงว่าแบบทดสอบนั้นมีแนวโน้มเป็นมิติเดียว

2. ความเป็นอิสระในการตอบข้อสอบ (Local Independence) หมายถึง โอกาสของการตอบข้อสอบแต่ละข้อของผู้สอบแต่ละคนมีความเป็นอิสระจากกัน ซึ่งประกอบด้วย

2.1 ความเป็นอิสระระหว่างข้อสอบ หมายความว่า ข้อสอบแต่ละข้อเป็นอิสระจากกัน กล่าวคือ การตอบข้อสอบข้อหนึ่งไม่มีผลต่อการตอบข้อสอบข้ออื่น ๆ ในแบบทดสอบฉบับนั้น นั่นคือ ถ้า $P_i(\theta)$ เป็นอิสระจาก $P_j(\theta)$ แล้ว $P_i(\theta) \cap P_j(\theta) = P_i(\theta) \cdot P_j(\theta)$

2.2 ความเป็นอิสระระหว่างผู้สอบ หมายความว่า ผู้สอบแต่ละคนตอบข้อสอบแต่ละข้ออย่างเป็นอิสระจากกัน กล่าวคือ ผลการตอบข้อสอบของผู้สอบคนหนึ่งไม่มีผลต่อการตอบข้อสอบของผู้สอบคนอื่น ๆ นั่นคือ ถ้า $P_i(\theta_A)$ เป็นอิสระจาก $P_i(\theta_B)$ แล้ว $P_i(\theta_A) \cap P_i(\theta_B) = P_i(\theta_A) \cdot P_i(\theta_B)$

การตรวจสอบความเป็นอิสระในการตอบข้อสอบสามารถทำได้โดย การทดสอบไค-สแควร์ ของข้อสอบแต่ละคู่ที่ระดับความสามารถเดียวกัน นอกจากนี้ ลอร์ด (Lord, 1980 : 19) กล่าวว่า “ความเป็นอิสระในการตอบข้อสอบเป็นคุณสมบัติที่เกิดขึ้นโดยอัตโนมัติ (automatically) จากคุณสมบัติของความเป็นมิติเดียว ไม่ได้เป็นข้อตกลงเบื้องต้นที่เพิ่มขึ้นมา” กล่าวคือ ถ้าแบบทดสอบฉบับใดมีความเป็นมิติเดียวแล้วแบบทดสอบฉบับนั้น จะมีความเป็นอิสระในการตอบข้อสอบโดยอัตโนมัติ

3. ฟังก์ชันการตอบข้อสอบ (Item Response Function : IRF) หรือโค้งคุณลักษณะข้อสอบ (Item Characteristic Curve : ICC) เป็นฟังก์ชันที่แสดงความสัมพันธ์ระหว่างระดับความสามารถของผู้สอบ (θ) กับโอกาสหรือความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง ; $P(\theta)$ แสดงโดยโค้งคุณลักษณะข้อสอบของแต่ละโมเดลที่ใช้ โดยที่รูปร่างของโค้งคุณลักษณะข้อสอบในแต่ละข้อมีคุณสมบัติไม่แปรเปลี่ยน (invariant) ไปตามกลุ่มผู้สอบ (Hambleton and Swaminathan, 1985 : 26)

4. ข้อสอบที่ใช้ต้องไม่เป็นข้อสอบประเภทความเร็ว (Speededness) ผู้สอบทุกคนควรมีโอกาสทำข้อสอบทุกข้อ เพื่อให้คะแนนรวมจากการทำแบบทดสอบกับลักษณะข้อสอบเป็นตัวประมาณค่าความสามารถที่แท้จริงของผู้สอบโดยไม่มีข้อจำกัดเกี่ยวกับปัจจัยเรื่องเวลาที่ใช้ในการสอบ (Hambleton & Swaminathan, 1985 : 30)

โมเดลในทฤษฎีการตอบข้อสอบ (Item Response Model) ได้รับการพัฒนาอย่างต่อเนื่องมานาน เช่น โมเดลนอร์มัลออร์จีฟ (Normal Ogive Model) โมเดลของราซช์ (Rasch Model) และโมเดลโลจิสติก (Logistic Model) เป็นต้น ซึ่งโมเดลดังกล่าว เรียกว่า dichotomous IRT Model เป็นโมเดลที่ใช้กับข้อสอบที่มีการตรวจให้คะแนนแบบ 2 ค่า (dichotomous) คือ ให้คะแนนเป็นแบบ 0-1 (ตอบถูกได้ 1 คะแนน, ตอบผิดได้ 0 คะแนน) มักนิยมใช้โมเดลโลจิสติกแบบ 1, 2 และ 3 พารามิเตอร์

โมเดลโลจิสติกแบบ 3 พารามิเตอร์ (3 PLM) เขียนในรูปคณิตศาสตร์ ได้ดังนี้

$$P_i(\theta) = C_i + (1 - C_i) \{1 + \exp[-Da_i(\theta - b_i)]\}^{-1}$$

เมื่อ $P_i(\theta)$ แทน โอกาสหรือความน่าจะเป็นของผู้สอบที่มีระดับความสามารถ θ จะตอบข้อสอบข้อที่ i ได้ถูกต้อง

θ แทน ค่าความสามารถของผู้สอบ

a_i แทน ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบข้อที่ i

b_i แทน ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i

c_i แทน ค่าพารามิเตอร์การเดาของข้อสอบข้อที่ i

ถ้ากำหนดให้พารามิเตอร์การเดา (c_i) มีค่าเป็นศูนย์ จะกลายเป็นโมเดลโลจิสติกแบบ 2 พารามิเตอร์ (2 PLM) ดังนี้

$$P_i(\theta) = \{1 + \exp[-Da_i(\theta - b_i)]\}^{-1}$$

โมเดลโลจิสติกแบบ 2 พารามิเตอร์นี้ ถ้ากำหนดให้พารามิเตอร์อำนาจจำแนก (a_i) มีค่าคงที่ทุกข้อ ($\bar{a}$) ในขณะที่พารามิเตอร์การเดามีค่าเป็นศูนย์ โมเดลนี้จะกลายเป็นโมเดลโลจิสติกแบบ 1 พารามิเตอร์ (1PLM) หรือ Rasch Model ซึ่งเป็นโมเดลที่อธิบายคุณลักษณะของข้อสอบด้วยพารามิเตอร์ความยาก (b_i) เพียงตัวเดียวเท่านั้น ดังนี้

$$P_i(\theta) = \{1 + \exp[-D\bar{a}(\theta - b_i)]\}^{-1}$$

เมื่อ D แทน ค่าองค์ประกอบของการปรับสเกล มีค่าเท่ากับ 1.7

$\exp$ แทน ค่าคงที่ของ natural logarithm มีค่าประมาณ 2.71818...

เมื่อใช้โมเดล IRT ที่เป็น dichotomous IRT model ดังกล่าวข้างต้นมาวิเคราะห์ข้อสอบแบบมาตราประมาณค่า แบบวัดเจตคติ หรือ แบบทดสอบวัดผลสัมฤทธิ์ ที่มีการตรวจให้คะแนนแบบหลายค่า (polytomous) สารสนเทศเกี่ยวกับการตอบข้อสอบส่วนหนึ่งจะถูกตัดทิ้งไป โดยถ้าเป็นการวิเคราะห์ข้อสอบแบบหลายตัวเลือกจะต้องกำหนดให้คะแนนสำหรับตัวถูกเป็น 1 ตัวเลือก นอกนั้นได้คะแนนเป็น 0 ทั้งหมด หรือถ้าวิเคราะห์ข้อสอบที่เป็นมาตราประมาณค่าอาจกำหนดให้การตอบในลำดับชั้น (category) ที่ 1, 2 และ 3 ได้คะแนนเป็น 0 ส่วนการตอบในลำดับชั้นที่ 4, 5, 6 และ 7 ได้คะแนนเป็น 1 เป็นต้น จึงจะทำการวิเคราะห์โดยใช้โมเดลการตอบข้อสอบแบบ 2 ค่าได้ ซึ่งทำให้สูญเสียสารสนเทศเกี่ยวกับแบบแผนการตอบส่วนหนึ่งไป (ศิริเดช สุชีวะ, 2539 : 17 อ้างจาก Muraki, 1992 ; Donoghue, 1994 ; Hambleton and Zaal, 1991 ; Drawsgow, F.1995)

ต่อมาได้มีการคิดค้นและพัฒนาโมเดลการตอบข้อสอบแบบหลายค่า (polytomous IRT model) ขึ้นมาอย่างต่อเนื่อง เพื่อช่วยแก้ปัญหาความไม่เหมาะสมระหว่างข้อมูลกับโมเดลการวิเคราะห์ โดยการตอบแต่ละตัวเลือกมีฟังก์ชันการตอบ (response function) ทำให้สารสนเทศเกี่ยวกับการตอบข้อสอบทั้งหมดยังคงอยู่ในการวิเคราะห์ตลอดกระบวนการ โมเดลของทฤษฎีการตอบข้อสอบที่ใช้กับแบบวัดและแบบทดสอบที่มีวิธีการตรวจให้คะแนนแบบหลายค่ามีอยู่หลายโมเดล โมเดลที่สำคัญ ได้แก่ Graded Response Model (GRM) พัฒนาขึ้นในปี 1960 โดย ซาโมจิมา (Samejima), Nominal Response Model (NRM) พัฒนาขึ้นในปี 1972 โดยบอค (Bock), Rating Scale Model (RSM) พัฒนาขึ้นในปี 1978 โดย แอนดริช (Andrich), Partial Credit Model (PCM) พัฒนาขึ้นในปี 1982 โดย มาสเตอร์ (Masters), Successive Interval Model (SIM) พัฒนาขึ้นในปี 1988 โดย รอสต์ (Rost) และ Generalized Partial Credit Model (GPCM) พัฒนามาจาก PCM โดย มูรากิ (Muraki) ในปี 1992

Generalized Partial Credit Model : GPCM

มูรากิ (Muraki, 1992) ได้พัฒนาโมเดลนี้จาก โมเดลพาเชียลเครดิต (Partial Credit Model : PCM) โดยปรับขยายโมเดลโลจิสติก 1 PLM ที่ว่าข้อสอบทุกข้อมีค่าอำนาจจำแนกเท่ากัน เป็นโมเดลโลจิสติกแบบ 2 PLM โดยรวมค่าอำนาจจำแนกหรือพารามิเตอร์ความชัน a_j ที่แปรเปลี่ยนแยกกันระหว่างข้อสอบที่ให้คะแนนแบบหลายค่าในแต่ละข้อ และใช้ EM algorithm มาร่วมกันประมาณค่าพารามิเตอร์โมเดล ด้วยวิธี marginal maximum likelihood โดยใช้โปรแกรม PARSCALE ในการประมาณค่าพารามิเตอร์โมเดล

GPCM ถูกสร้างขึ้นจากพื้นฐานของข้อตกลงเบื้องต้น ที่ว่า ในการตอบข้อสอบข้อที่มีคำตอบหลายลำดับชั้น (Multicategory) บุคคลที่มีความสามารถหรือมีคุณลักษณะที่ต้องการวัดสูง มีความน่าจะเป็นในการเลือกตอบลำดับชั้นการตอบ (response category) ที่ k มากกว่าลำดับชั้นคำตอบที่ $k-1$ นั่นคือ

$$P_{jk|k-1,k}(\theta) = \frac{P_{jk}(\theta)}{P_{j,k-1}(\theta) + P_{jk}(\theta)} = \frac{\exp[D a_j(\theta - b_{jk})]}{1 + \exp[D a_j(\theta - b_{jk})]} \quad (12)$$

เมื่อ $k = 2, 3, \dots, m_j$ และ m_j เป็นจำนวนของลำดับชั้นการตอบ

ถ้าให้ $P_{jk}(\theta)$ แทน โอกาสหรือความน่าจะเป็นของการเลือกตอบลำดับชั้นที่ k จาก m_j ลำดับชั้นที่เป็นไปได้ของข้อสอบข้อที่ j

ดังนั้น GPCM เขียนได้เป็น

$$P_{jk}(\theta) = \frac{\exp\left[\sum_{v=1}^k Z_{jv}(\theta)\right]}{\sum_{c=1}^{m_j} \exp\left[\sum_{v=1}^c Z_{jv}(\theta)\right]} \quad (13)$$

$$\text{และ } Z_{jv}(\theta) = D a_j(\theta - b_{jv}) = D a_j(\theta - b_j + d_v)$$

เมื่อ D เป็น ค่าคงที่ของสเกลที่จัดให้สเกลคุณลักษณะ (θ) ที่วัดได้เหมือนกับ Normal ogive model ($D = 1.7$)

a_j เป็น พารามิเตอร์อำนาจจำแนกหรือพารามิเตอร์ความชัน (slope parameter)

b_{jv} เป็น พารามิเตอร์ลำดับชั้นข้อสอบ (item – category parameter)

b_j เป็น พารามิเตอร์ตำแหน่งข้อสอบ (item location parameter)

d_v เป็น พารามิเตอร์ลำดับชั้น (category parameter)

โดยที่ $b_{jv} = b_j - d_v$
นั่นคือ

$$P_{jk}(\theta) = \frac{\exp \left[\sum_{v=1}^k D a_j (\theta - b_{jv}) \right]}{\sum_{c=1}^{m_j} \exp \left[\sum_{v=1}^c D a_j (\theta - b_{jv}) \right]} \quad (14)$$

มาสเตอร์ (Muraki. 1992 : 162 citing Master. 1982) เรียก พารามิเตอร์ลำดับชั้นข้อสอบ (item step parameter); b_{jk} ว่าเป็นเหมือนพารามิเตอร์ความยากของข้อสอบ (item difficulty parameter) ในแต่ละลำดับชั้น ซึ่งเป็นจุดบนสเกลความสามารถ θ ณ ที่ $P_{j,k-1}(\theta)$ กับ $P_{jk}(\theta)$ ตัดกัน ซึ่งเส้นโค้งทั้ง 2 เส้นถือเป็นฟังก์ชันการตอบลำดับชั้นข้อสอบ (item category response function : ICRF_s) ใน 2 ลำดับค่าคะแนนที่ตัดกันเพียงจุดเดียว บนค่าใด ๆ ตลอดช่วงสเกล θ นั่นคือ

ถ้า $\theta = b_{jk}$, $P_{jk}(\theta) = P_{j,k-1}(\theta)$

ถ้า $\theta > b_{jk}$, $P_{jk}(\theta) > P_{j,k-1}(\theta)$

ถ้า $\theta < b_{jk}$, $P_{jk}(\theta) < P_{j,k-1}(\theta)$

ภายใต้ข้อตกลงที่ว่า $a_j > 0$ และข้อสอบข้อที่ j นั้น b_{jk} ไม่จำเป็นต้องจัดเรียงลำดับ เพราะค่าพารามิเตอร์จะแสดงถึงขนาดที่สัมพันธ์กันของความน่าจะเป็นในการตอบ $P_{j,k-1}(\theta)$ และ $P_{jk}(\theta)$ ที่อยู่ในลำดับติดกัน ส่วนพารามิเตอร์ a_j เป็นพารามิเตอร์ความชันของข้อสอบข้อที่ j โดยปกติค่า a_j มีพิสัยที่เป็นไปได้ในช่วง 0 ถึง ∞ ซึ่งค่าพารามิเตอร์ความชันจะบ่งชี้ระดับ (degree) ที่การตอบในแต่ละลำดับชั้นแปรเปลี่ยนไปในแต่ละข้อ เช่นเดียวกับระดับความสามารถ θ ที่แปรเปลี่ยนไป

ความต่อเนื่องของโมเดล 2 พารามิเตอร์ของ เบอร์นบอม (Birnbbaum. 1968) จากการณีการตอบแบบ dichotomous ไปสู่กรณีการตอบแบบ polytomous มีการเชื่อมโยงแนวคิดของการทดสอบตามแบบมาตรฐานเดิมเป็นอย่างดีด้วยการยังคงเก็บค่าอำนาจจำแนกของข้อสอบไว้ในโมเดล โมเดลที่มีค่าพารามิเตอร์ความชันสามารถขยายไปสู่รูปแบบที่เป็นหลายมิติ (multidimensional form)(Muraki. 1985, 1990b, 1992 : 162)

สารสนเทศของข้อสอบ (Item Information)

สำหรับโมเดลการตอบข้อสอบที่ให้คะแนนแบบหลายค่า (Polytomous IRT Model) ของ GPCM มีดังนี้

1. ฟังก์ชันสารสนเทศของข้อสอบ (Item Information Function : IIF)

เขียนแทนด้วยสมการ

$$I_j(\theta) = D^2 a_j^2 \left[\sum_{k=0}^{m_j} k^2 P_{jk}(\theta) - \left(\sum_{k=0}^{m_j} k P_{jk}(\theta) \right)^2 \right] \quad (15)$$

เมื่อ $I_j(\theta)$ คือ ฟังก์ชันสารสนเทศของข้อสอบข้อที่ j ในช่วงพิสัยความสามารถ θ
 D คือ ค่าคงที่ของสเกลมีค่าประมาณ 1.7
 a_j คือ ค่าพารามิเตอร์อำนาจจำแนก(ความชัน) ของข้อสอบ
 k คือ ค่าคะแนนลำดับชั้น (category scores)
 $P_{jk}(\theta)$ คือ ความน่าจะเป็นของผู้สอบที่มีความสามารถ θ ในการตอบข้อสอบข้อที่ j ได้คะแนน k

2. ฟังก์ชันสารสนเทศลำดับชั้นข้อสอบ (Item Category Information Function : ICIF) บอกรัก (Bock, 1972) ได้เสนอ การนิยามสารสนเทศที่เกิดจากการตอบในลำดับที่ k ของข้อสอบข้อที่ j ว่าเป็นการแบ่งส่วน (partition) ของสารสนเทศข้อสอบที่เกิดจากลำดับชั้นนั้น นั่นคือ

$$I_{jk}(\theta) = P_{jk}(\theta) I_j(\theta) \quad (16)$$

เรียกว่า ฟังก์ชันสารสนเทศลำดับชั้นข้อสอบ (ICIF) ซึ่ง ICIF ของ $I_{j,k-1}(\theta)$ กับ ICIF ของ $I_{jk}(\theta)$ จะตัดกันที่จุด b_{jk} บนสเกล θ
 ทั้งนี้ IIF สามารถแสดงเป็น ผลรวมของ ICIF ได้ดังนี้

$$I_j(\theta) = \sum_{k=1}^{m_j} I_{jk}(\theta) \quad (17)$$

3. ฟังก์ชันสารสนเทศของแบบทดสอบ (Test Information Function : TIF) ถูกกำหนดเป็นผลรวมของ IIF ได้ดังนี้

$$I(\theta) = \sum_{j=1}^n I_j(\theta) \quad (18)$$

การประมาณค่าพารามิเตอร์ (Parameter Estimation)

ให้ U_{jki} เป็นสมาชิกในเมตริกซ์ของรูปแบบการตอบที่สังเกตได้ i

ถ้าการตอบข้อสอบข้อที่ j อยู่ในลำดับชั้นที่ k แล้ว $U_{jki} = 1$ นอกนั้น $U_{jki} = 0$

ความน่าจะเป็นอย่างมีเงื่อนไข (conditional probability) ของรูปแบบการตอบ i ในการทำข้อสอบจำนวน n ข้อที่มี m ลำดับชั้นการตอบ ด้วยความสามารถ θ ซึ่งตามหลักของ local independence นั้น เมื่อแทนด้วยเมตริกซ์การตอบ $[U_i = (U_{jk})_i]$ ความน่าจะเป็นร่วม (Joint probability) ก็คือ

$$P(U_i | \theta) = P_i[(U_{jk}) | \theta] = \prod_{j=1}^n \prod_{k=1}^m [P_{jk}(\theta)]^{U_{jki}} \quad (19)$$

สำหรับกลุ่มผู้สอบที่เลือกมาอย่างสุ่มจากประชากรที่มีการแจกแจงปกติของตัวแปรคุณลักษณะแฝงภายใน, $\phi(\theta)$ นั้น ความน่าจะเป็นแบบข้างขอบ (marginal probability) ของรูปแบบการตอบที่สังเกตได้ i คือ

$$P_i[(U_{jk})] = \int_{-\infty}^{\infty} P_i[(U_{jk})|\theta] \phi(\theta) d\theta \quad (20)$$

ถ้าผู้สอบสามารถตอบข้อสอบได้ถึงข้อที่ n ซึ่งมี m ลำดับชั้นการตอบของผู้สอบ รูปแบบ i สามารถกำหนดเป็น 1 รูปแบบ ใน m^n รูปแบบการตอบที่เป็นอิสระจากกัน (m^n mutually exclusive pattern)

ให้ r_i แทน จำนวนผู้สอบที่ตอบในรูปแบบ i

และ N แทน จำนวนผู้สอบทั้งหมดที่สุ่มจากประชากร

แล้ว r_i มีการแจกแจงแบบ multinomially ด้วยพารามิเตอร์ N และ $P_i[(U_{jk})]$

ดังนั้น

$$L = [N! / \prod_{i=1}^{m^n} r_i] \prod_{i=1}^{m^n} \{P_i[(U_{jk})]\}^{r_i} \quad (21)$$

ใส่ natural logarithm ในสมการ 21 จะได้

$$\ln L = \ln N! - \sum_{i=1}^{m^n} \ln r_i! + \sum_{i=1}^{m^n} r_i \ln P_i[(U_{jk})] \quad (22)$$

การประมาณค่าพารามิเตอร์ข้อสอบ (Item parameter estimation)

ให้ v_j แทน พารามิเตอร์ a_j หรือ b_j

v_h แทน พารามิเตอร์ v_j ของข้อสอบข้อที่ $j = h$

ความน่าจะเป็น (likelihood) ของสมการ 22 สามารถหาอนุพันธ์ได้จาก

$$\frac{\partial \ln L}{\partial v_h} = \sum_{i=1}^{m^n} \frac{r_i}{P_i[(U_{jk})]} \int_{-\infty}^{\infty} P_i[(U_{jk})|\theta] \sum_{k=1}^m \frac{\partial [P_{hk}(\theta)]^{v_{hki}}}{\partial v_h} \frac{\phi(\theta) d\theta}{[P_{hk}(\theta)]^{v_{hki}}} \quad (23)$$

รูปแบบของคะแนนที่สังเกตได้ กำหนดด้วย $\ell = 1, 2, \dots, S$ เมื่อ $S \leq \min(N, m^n)$

r_ℓ แทน จำนวนผู้สอบมีรูปแบบการตอบ ℓ แล้ว

$$\sum_{\ell=1}^S r_\ell = N \quad (24)$$

อนุพันธ์ลำดับแรก ของฟังก์ชันความน่าจะเป็น (likelihood function) ในสมการ 23 สามารถประมาณค่าได้โดยใช้ Gauss – Hermite quadrature ดังนั้น

$$\frac{\partial \ln L}{\partial v_h} \approx \sum_{\ell=1}^S \sum_{f=1}^F \frac{r_\ell L_\ell(x_f) A(x_f)}{\tilde{p}_\ell} \sum_{k=1}^m \frac{\partial [P_{hk}(x_f)]^{v_{hkf}}}{\partial v_h} \frac{1}{[P_{hk}(x_f)]^{v_{hkf}}} \quad (25)$$

$$\text{เมื่อ } \tilde{p}_\ell = \sum_{f=1}^F L_\ell(x_f) A(x_f)$$

$$\text{และ } L_\ell(x_f) = \prod_{j=1}^n \prod_{k=1}^m [p_{jk}(x_f)]^{u_{jkl}}$$

จากสมการ 25 ; x_f เป็น quadrature point

และ $A(x_f)$ คือ ค่าน้ำหนักของ Gauss – Hermite quadrature ซึ่งเป็นความหนาแน่นของความน่าจะเป็นปกติมาตรฐาน (standard normal probability density) โดยประมาณ ณ จุด x_f ดังนั้น

$$\sum_{f=1}^F A(x_f) = 1 \quad \text{เมื่อ } F \text{ เป็นจำนวน quadrature point ทั้งหมด}$$

เนื่องจาก u_{hkf} สามารถหาค่าที่เป็นไปได้เพียง 2 ค่า คือ 1 หรือ 0 สมการ 25 เขียนใหม่เป็น

$$\sum_{f=1}^F \sum_{k=1}^m \frac{\bar{r}_{hkf}}{P_{hk}(x_f)} \frac{\partial P_{hk}(x_f)}{\partial v_h} \quad (26)$$

$$\text{เมื่อ } \bar{r}_{hkf} = \sum_{\ell=1}^S \frac{r_\ell L_\ell(x_f) A(x_f) u_{hkf}}{\tilde{p}_\ell} \quad (27)$$

และ $\bar{r}_{hkf}$ เป็นความถี่ที่คาดหวังชั่วคราว (provisional expected frequency) ของการตอบลำดับขั้นที่ k ของข้อสอบข้อที่ h ณ quadrature point ที่ f

บ็อกซ์และอิตกิน (Bock & Aitkin, 1981) ได้ใช้ EM algorithm ในการประมาณค่าพารามิเตอร์สำหรับข้อสอบแต่ละข้อโดยลำดับและกระบวนการทำซ้ำ (iteration process) จะกระทำซ้ำกับข้อสอบทั้งหมด n ข้อจนกระทั่งค่าประมาณของข้อสอบทั้งหมดมีความคงที่ตามที่กำหนด ซึ่งจำนวนรอบของกระบวนการทำซ้ำที่ q (q^{th} cycle of the iteration process) สามารถแสดงได้เป็น

$$v_q = v_{q-1} + V^{-1} t \quad (28)$$

เมื่อ v_q และ v_{q-1} เป็นค่าประมาณพารามิเตอร์ของการทำซ้ำรอบที่ q และ q-1 ตามลำดับ

V^{-1} เป็นอินเวอร์สของเมตริกซ์สารสนเทศ

และ t เป็น gradient vector

สมาชิกของ t และ V ในการประมาณค่าพารามิเตอร์ข้อสอบ คือ

$$t_{vh} = \sum_{f=1}^F \sum_{k=1}^m \frac{\bar{r}_{hkf}}{P_{hk}(x_f)} \frac{\partial P_{hk}(x_f)}{\partial v_h} \quad (29)$$

$$\text{และ } V_{v_h v_h} = \sum_{f=1}^F \sum_{k=1}^m \frac{1}{P_{hk}(x_f)} \frac{\partial P_{hk}(x_f)}{\partial v_h} \frac{\partial P_{hk}(x_f)}{\partial v_h} \quad (30)$$

$$\begin{aligned} \text{เมื่อ } u_h &= a_h \text{ หรือ } b_h \\ \omega_h &= a_h \text{ หรือ } b_h \end{aligned}$$

ส่วน $\bar{N}_f$ ในสมการ 30 เรียกว่า ขนาดของกลุ่มตัวอย่างที่คาดหวังชั่วคราว (provisional expected sample size) ณ quadrature point ที่ f และคำนวณได้โดย

$$\bar{N}_f = \sum_{\ell=1}^S \frac{r_\ell L_\ell(x_f) A(x_f)}{\tilde{P}_\ell} \quad (31)$$

ส่วน ความน่าจะเป็นสูงสุดแบบมาร์จินอลใน EM algorithm (The Marginal Maximum Likelihood (MML) EM algorithm) ที่ใช้ในโปรแกรม PARSCALE ประกอบด้วย 2 ขั้นตอน

ขั้นตอนแรก เป็นขั้นการคาดหวัง (Expectation Step : E-step) ซึ่งความถี่ที่คาดหวังชั่วคราว และขนาดของกลุ่มตัวอย่างที่คาดหวังชั่วคราว ถูกนำมาคำนวณด้วยสมการ 27 และ 31 ตามลำดับ

ขั้นตอนที่สอง เป็นขั้นการประมาณค่าสูงสุด (Maximization Step : M-step) ค่าประมาณความน่าจะเป็นสูงสุดแบบมาร์จินอล (MML) หาได้ด้วยสมการ 28

ทั้ง E-step และ M-step จะถูกกระทำซ้ำๆ (EM cycle) จนกระทั่ง ค่าประมาณทั้งหมดคงที่

ในการศึกษาการทำหน้าที่ต่างกันของข้อสอบตามทฤษฎี IRT ค่าประมาณพารามิเตอร์ของข้อสอบที่ได้จากข้อสอบข้อเดียวกัน แต่วิเคราะห์จากผู้สอบที่อยู่ต่างกลุ่มกัน ไม่สามารถนำมาเปรียบเทียบกันได้ เนื่องจากข้อตกลงเบื้องต้นของทฤษฎี IRT ที่ว่า ค่าพารามิเตอร์ของข้อสอบจะต้องไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบ ดังนั้น ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ จะต้องนำค่าพารามิเตอร์ของข้อสอบมาเปรียบเทียบ (equating) ให้อยู่บนสเกลเมตริกซ์เดียวกัน วิธีการที่นิยมใช้ คือ การแปลงค่าเชิงเส้น (linear transformation) โดยที่ค่าประมาณจากการเปรียบเทียบของกลุ่มสนใจ จะแปลงให้อยู่บนสเกลเมตริกซ์ของกลุ่มอ้างอิง ดังนี้ (Kim & Cohen, 1992 : 56)

$$\begin{aligned} a_{jF}^* &= a_{jF} / A \\ b_{jF}^* &= A b_{jF} + K \end{aligned}$$

- เมื่อ a_{jF}^* แทน ค่าอำนาจจำแนกของข้อสอบข้อที่ j จากผู้สอบกลุ่มสนใจที่แปลงค่าแล้ว
 b_{jF}^* แทน ค่าความยากของข้อสอบข้อที่ j จากผู้สอบกลุ่มสนใจที่แปลงค่าแล้ว
 a_{jF} แทน ค่าอำนาจจำแนกของข้อสอบข้อที่ j จากผู้สอบกลุ่มสนใจที่ยังไม่ได้แปลงค่า
 b_{jF} แทน ค่าความยากของข้อสอบข้อที่ j จากผู้สอบกลุ่มสนใจที่ยังไม่ได้แปลงค่า
 A แทน ค่าความชัน (slope)
 K แทน ค่าจุดตัด (Intercept)

ส่วนค่าประมาณพารามิเตอร์ความสามารถของผู้สอบ i ของกลุ่มสนใจ สามารถแปลงค่าให้อยู่บนสเกลของกลุ่มอ้างอิง ได้ดังนี้

$$\theta_{IF}^* = A \theta_{IF} + K$$

เมื่อปรับเทียบค่าพารามิเตอร์ของข้อสอบให้อยู่บนสเกลเดียวกันแล้วจึงวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) ตามทฤษฎีการตอบข้อสอบ คือ การที่ฟังก์ชันการตอบข้อสอบข้อนั้นของประชากรย่อย 2 กลุ่ม มีความแตกต่างกัน ทำให้เกิดพื้นที่ระหว่างโค้งคุณลักษณะข้อสอบ ดังนั้น การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบตามทฤษฎีนี้ สามารถกระทำได้โดยการเปรียบเทียบฟังก์ชันการตอบข้อสอบของผู้สอบกลุ่มย่อยที่ต้องการศึกษา ซึ่งแบ่งออกเป็น 2 วิธี ดังนี้ (Hambleton, Swaminathan and Rogers. 1991 : 110; Kim and Cohen. 1995 : 291)

1. วิธีการวัดพื้นที่ (Area Measures)

เป็นวิธีการเปรียบเทียบผลการตอบข้อสอบ โดยการคำนวณค่าประมาณพื้นที่ระหว่างโค้งคุณลักษณะข้อสอบจากผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจที่มีความสามารถระดับเดียวกัน ซึ่งมีสูตรทั่วไปที่ใช้ในการคำนวณพื้นที่ ดังนี้ (Millsap and Everson. 1993 : 310)

$$A = f_s [P_R(\theta) - P_F(\theta)]$$

เมื่อ A แทน ดัชนีพื้นที่

f_s แทน ฟังก์ชัน f ในช่วง s ซึ่ง $s = (\theta_L, \theta_H)$ โดยที่ L และ H เป็นค่าต่ำสุดและสูงสุด ตามลำดับ

$P_R(\theta)$ แทน โอกาสของการตอบข้อสอบถูกต้องที่ระดับความสามารถ θ ของผู้สอบกลุ่มอ้างอิง

$P_F(\theta)$ แทน โอกาสของการตอบข้อสอบถูกต้องที่ระดับความสามารถ θ ของผู้สอบกลุ่มสนใจ

วิธีการนี้มีผู้ศึกษาค้นคว้าไว้หลายคน เช่น รัดเนอร์ (Rudner, 1977) ราชู (Raju, 1988, 1990), ลินน์, เลวิน, ฮาสติงส์และวาร์ลดรอป (Linn, Levine, Hastings and Wardrop, 1981), คิมและโคเฮน (Kim and Cohen, 1991) และ วายเนอร์ (Wainer, 1993) เป็นต้น แนวคิดที่ได้รับความนิยม ได้แก่ แนวคิดของราชู และ แนวคิดของ คิมและโคเฮน

1.1 การวัดพื้นที่ตามแนวคิดของราชู

ราชู (Raju, 1990) ได้เสนอสูตรการวัดพื้นที่บนช่วงความสามารถเปิด (Open Interval Area) บางครั้งเรียกว่า การวัดพื้นที่สมบูรณ์ (Exact Area Measures) ซึ่งในการคำนวณพื้นที่สามารถแบ่งเป็น 2 ลักษณะ คือ

1.1.1 การคำนวณพื้นที่แบบคิดเครื่องหมายบนช่วงเปิด (Open Interval Signed Area หรือ Exact Signes Area : ESA) ซึ่งคำนวณได้จากสูตร

$$ESA = \int_{-\infty}^{+\infty} [P_R(\theta) - P_F(\theta)] d\theta$$

1.1.2 การคำนวณพื้นที่แบบไม่คิดเครื่องหมายบนช่วงเปิด (Open Interval Unsigned Area หรือ Exact Unsignes Area : EUA) ซึ่งคำนวณได้จากสูตร

$$EUA = \int_{-\infty}^{+\infty} |P_R(\theta) - P_F(\theta)| d\theta$$

ในการตัดสินใจว่าข้อสอบทำหน้าที่ต่างกันหรือไม่ ราชู ได้เสนอการนำดัชนีการวัดพื้นที่มาใช้ทดสอบนัยสำคัญ โดยใช้สถิติ Z ภายใต้สมมติฐานของการแจกแจงปกติ ซึ่งแบ่งเป็น 2 ชนิด คือ

ชนิดที่ 1 การทดสอบนัยสำคัญของ ESA ใช้สถิติทดสอบ Z จากสูตร

$$Z_i(ESA) = \frac{\hat{b}_{iF} - \hat{b}_{iR}}{\sqrt{\text{Var}(\hat{b}_{iF}) + \text{Var}(\hat{b}_{iR})}}$$

โดยนำดัชนี ESA ของข้อสอบข้อที่ i ทดสอบความแตกต่างกับ 0

ชนิดที่ 2 การทดสอบนัยสำคัญของ EUA สามารถแบ่งออกเป็น 2 กรณี คือ

กรณีที่ 1 เมื่อ $\hat{a}_{iR} = \hat{a}_{iF}$ การทดสอบเหมือนกับการทดสอบดัชนี ESA

กรณีที่ 2 เมื่อ $\hat{a}_{iR} \neq \hat{a}_{iF}$ ใช้สถิติทดสอบ Z จากสูตร

$$Z_i(H) = \frac{H_i}{\sqrt{\text{Var}(H_i)}}$$

โดยนำดัชนี EUA ของข้อสอบข้อที่ i ไปทดสอบความแตกต่างกับ 0

1.2 การวัดพื้นที่ตามแนวคิดของ คิม และ โคเฮน

คิม และ โคเฮน (Kim and Cohen, 1991) ได้เสนอสูตรการวัดพื้นที่บนช่วงความสามารถปิด (Closed Interval Area) โดยคำนวณบนช่วงความสามารถ $[\theta_1, \theta_2]$ ในการคำนวณพื้นที่ที่สามารถแบ่งเป็น 2 ลักษณะ คือ

1.2.1 การคำนวณพื้นที่แบบคิดเครื่องหมายในช่วงปิด (Closed Interval Signed Area : CSA) ซึ่งคำนวณได้จากสูตร

$$CSA = \int_{\theta_1}^{\theta_2} [P_R(\theta) - P_F(\theta)] d\theta = S_R(\theta_1, \theta_2) - S_F(\theta_1, \theta_2)$$

1.2.2 การคำนวณพื้นที่แบบไม่คิดเครื่องหมายในช่วงปิด (Closed Interval Unsigned Area : CUA) ซึ่งคำนวณได้จากสูตร

$$CSA = \int_{\theta_1}^{\theta_2} |P_R(\theta) - P_F(\theta)| d\theta = |S_R(\theta_1, \theta_2) - S_F(\theta_1, \theta_2)|$$

ในการตัดสินใจการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีการวัดพื้นที่ของคิมและโคเฮนจะไม่ใช้การทดสอบนัยสำคัญ แต่ใช้วิธีการเปรียบเทียบขนาดของพื้นที่กับเกณฑ์ที่กำหนดขึ้น

2. วิธีเปรียบเทียบค่าพารามิเตอร์ของข้อสอบ

เป็นวิธีการเปรียบเทียบโดยการนำค่าพารามิเตอร์ เช่นค่าอำนาจจำแนก ค่าความยากของข้อสอบระหว่างผู้สอบกลุ่มย่อย 2 กลุ่มมาเปรียบเทียบกันแล้วทดสอบนัยสำคัญว่ามีความแตกต่างกันหรือไม่ วิธีการนี้

มีผู้ศึกษาหลายคน ได้แก่ วิธีเปลี่ยนค่าความยาก (Difficulty Shift) ของ ไรท์ และ สโตน (Wright and Stone, 1979) วิธีการทดสอบไค-สแควร์ของลอร์ด (Lord's χ^2) (Lord, 1980) วิธีการตอบข้อสอบแบบเทียม (pseudo-IRT) ของลินน์และฮาร์นิช (Linn and Harnisch, 1981) วิธีการทดสอบ F ของฮูลิน, คลาสโก และ โคโมคาร์ (Hulin, Drasgow and Komocar, 1982) และวิธีการทดสอบอัตราส่วนความน่าจะเป็น (Likelihood ratio test : LR) ของ ทิสเซน สเตอ์นเบิร์ก และวายนอร์ (Thissen, Steinberg and Wainer, 1993) แต่วิธีการที่ได้รับความนิยมมาก ได้แก่วิธีการทดสอบ ไค-สแควร์ของลอร์ด (Lord's Chi-square test) ซึ่งวิธีการนี้สามารถใช้ได้กับโมเดลโลจิสติก แบบ 1, 2 และ 3 พารามิเตอร์ และมีประสิทธิภาพในการตรวจสอบ เมื่ออยู่บนพื้นฐานของสมมติฐาน 3 ประการ คือ

1. การทดสอบทางสถิติมีการแจกแจงไค-สแควร์แบบเชิงเส้นกำกับ (Asymptotic)
2. ต้องรู้ค่าความสามารถของผู้สอบ
3. ต้องใช้วิธีการประมาณค่าแบบความน่าจะเป็นสูงสุด (maximum likelihood estimate)

ในกรณี 2 พารามิเตอร์ การเปรียบเทียบค่าพารามิเตอร์ของข้อสอบจะทดสอบความแตกต่างเฉพาะพารามิเตอร์ความยาก (b) และอำนาจจำแนก (a) เท่านั้น โดยกำหนดให้พารามิเตอร์การเดา (c) มีค่าเป็นศูนย์ พอสรุปได้ดังนี้

1. ประมาณค่าพารามิเตอร์ของข้อสอบ
2. แปลงค่าพารามิเตอร์ให้อยู่บนสเกลหรือหน่วยเดียวกัน
3. นำค่าพารามิเตอร์ของข้อสอบมาเปรียบเทียบกัน โดยใช้สถิติทดสอบ χ^2

กำหนดสมมติฐานในการทดสอบ ดังนี้

$$\begin{aligned} H_0 : b_{iFm} &= b_{iRm} & \text{และ} & a_{iF} = a_{iR} \\ H_1 : b_{iFm} &\neq b_{iRm} & \text{และ} & a_{iF} \neq a_{iR} \end{aligned}$$

สถิติทดสอบ ไค-สแควร์ (χ^2) คือ

$$\chi_i^2 = V_i' \Gamma_i^{-1} V_i$$

โดยที่

$$V_i' = [\hat{a}_{iF} - \hat{a}_{iR}, \hat{b}_{iF1} - \hat{b}_{iR1}, \dots, \hat{b}_{iFm} - \hat{b}_{iRm}]$$

$$\Gamma_i = \Gamma_{iF} + \Gamma_{iR}$$

เมื่อ $\hat{a}_{iF}$ แทน ค่าประมาณพารามิเตอร์อำนาจจำแนกของข้อสอบข้อที่ i จากผู้สอบกลุ่มสนใจ
 $\hat{a}_{iR}$ แทน ค่าประมาณพารามิเตอร์อำนาจจำแนกของข้อสอบข้อที่ i จากผู้สอบกลุ่มอ้างอิง

$\hat{b}_{iF_m}$ แทน ค่าประมาณพารามิเตอร์ความยากของข้อสอบข้อที่ i จากผู้สอบกลุ่มสนใจ
ในลำดับขั้นที่ m

$\hat{b}_{iR_m}$ แทน ค่าประมาณพารามิเตอร์ความยากของข้อสอบข้อที่ i จากผู้สอบกลุ่มอ้างอิง
ในลำดับขั้นที่ m

V_i แทน เวกเตอร์ความแตกต่างของค่าประมาณพารามิเตอร์ของข้อสอบข้อที่ i
จากผู้สอบกลุ่มสนใจและกลุ่มอ้างอิง

Γ_i แทน เมตริกซ์ความแปรปรวน-ความแปรปรวนร่วมของ $(\hat{a}_{iF} - \hat{a}_{iR})$ และ
 $(\hat{b}_{iF_m} - \hat{b}_{iR_m})$ ที่มีขนาด $m \times m$

Γ_i^{-1} แทน อินเวอร์สของเมตริกซ์ Γ_i

ถ้าผลการทดสอบ พบว่า ค่าพารามิเตอร์แตกต่างอย่างมีนัยสำคัญ แสดงว่า ข้อสอบทำหน้าที่ต่างกัน

วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในแบบทดสอบที่มีการให้คะแนนแบบหลายค่าที่
ผู้วิจัยสนใจศึกษา คือ วิธี GMH และวิธี Polytomous SIBTEST รวมทั้งวิธี GPCM ซึ่งผู้วิจัยนำมาใช้เป็น
เกณฑ์ในการตัดสินความถูกต้องของการตรวจสอบ นั้น แต่ละวิธีมีคุณสมบัติ มีข้อตกลงเบื้องต้น มีรูปแบบ
ของโมเดลและอีกหลายประเด็น รวมทั้งมีจุดเด่น จุดด้อยที่เหมือนกันและต่างกัน พอสรุปได้ดังตาราง 5

ตาราง 5 เปรียบเทียบวิธีการในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนแบบหลายค่า

ประเด็น	GMH	Polytomous SIBTEST	GPCM
ข้อตกลงเบื้องต้น	คะแนนรวมจากแบบทดสอบ เป็นตัวแทนความสามารถของผู้สอบ	คะแนนรวมจากแบบทดสอบ เป็นตัวแทนความสามารถของผู้สอบและมีมิติการวัด 2 มิติ คือคุณลักษณะแฝงที่ต้องการวัดและคุณลักษณะแฝงที่ไม่ต้องการวัด	แบบทดสอบเป็นมิติเดียว และโครงสร้างคุณลักษณะข้อสอบสามารถแสดงฟังก์ชันของค่าความสามารถและโอกาสในการตอบข้อสอบ
ทฤษฎีพื้นฐาน	Non-IRT	IRT ชนิดหลายมิติ	IRT
สิ่งที่ทำการวิเคราะห์	ความแตกต่างของสัดส่วนการตอบระหว่างผู้ที่มีความสามารถระดับเดียวกัน	ความแตกต่างระหว่างคะแนนเฉลี่ยและสัดส่วนการตอบข้อสอบระหว่างผู้สอบที่มีความสามารถระดับเดียวกัน	ความแตกต่างของฟังก์ชันการตอบข้อสอบของผู้สอบที่มีความสามารถระดับเดียวกัน
สิ่งที่ใช้ในการวิเคราะห์	ค่าดัชนี GMH^2 และ ความมีนัยสำคัญทางสถิติ	ค่าดัชนี β^* และความมีนัยสำคัญทางสถิติ	ความแตกต่างของค่าพารามิเตอร์และความมีนัยสำคัญ
รูปแบบของโมเดล	Non-Parametric	Non-Parametric	Non-Parametric
จุดเด่น*	คำนวณง่าย ใช้กลุ่มตัวอย่างน้อย เสียค่าใช้จ่ายน้อย สะดวกในทางปฏิบัติ	คำนวณง่าย ไม่ซับซ้อน ใช้กลุ่มตัวอย่างน้อย ตรวจสอบ DIF ได้ทั้งเป็นรายข้อและกลุ่มข้อสอบ วิเคราะห์ได้ทั้งแบบมิติเดียวและหลายมิติ	ค่าพารามิเตอร์ของข้อสอบไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบ มีความไวและเฉียบคมในการตรวจสอบ DIF ให้รายละเอียดมาก
จุดด้อย*	ไม่มีความไวในการตรวจสอบ DIF โดยเฉพาะ DIF แบบไม่สม่ำเสมอ	มีความผิดพลาดประเภทที่ 1 สูง เมื่อกลุ่มตัวอย่างมีการแจกแจงความสามารถแตกต่างกัน	การคำนวณยุ่งยาก ซับซ้อน หลายขั้นตอน ใช้กลุ่มตัวอย่างขนาดใหญ่ เสียค่าใช้จ่ายสูง

* สรุปรจากพื้นฐานของวิธี MH วิธี SIBTEST และวิธีการตามทฤษฎีการตอบข้อสอบ (IRT)

5. เอกสารและงานวิจัยที่เกี่ยวข้อง

งานวิจัยต่างประเทศ

สวิกค์ และคณะ (Zwick, et al. 1993 : 233-251) ได้ศึกษาวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่เป็นภาคขยายของวิธีการ Mantel-Haenszel (MH;1959) คือวิธี Mantel ที่ใช้กับข้อสอบที่ลำดับชั้นการตอบมีการจัดเรียง ซึ่งสถิติทดสอบที่ใช้คือ Mantel χ^2 กับวิธี Mantel ที่ใช้กับข้อสอบที่ลำดับชั้นการตอบไม่จำเป็นต้องจัดเรียง ซึ่งใช้สถิติทดสอบที่ใช้คือ $GMH\chi^2$ รวมทั้งศึกษาวิธีการที่ใช้สถิติเชิงพรรณนาที่เรียกว่า Standardized Mean Difference (SMD) เพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่เป็นการทดสอบการปฏิบัติที่มีการให้คะแนนแบบหลายค่า โดยศึกษาจากแบบทดสอบที่จำลองข้อมูลขึ้นจำนวน 25 ข้อ โดยมีข้อสอบ 24 ข้อ เป็นข้อสอบหลักที่ใช้เป็นตัวแปรจับคู่ ส่วนข้อที่ 25 เป็นข้อสอบที่ใช้ศึกษา เงื่อนไขในการจำลองข้อมูล คือ การแจกแจงความสามารถของกลุ่มสนใจมี 2 ระดับและคุณลักษณะของข้อสอบที่ศึกษามี 27 ระดับ โดยคุณลักษณะของข้อสอบนั้นประกอบด้วยข้อสอบที่มีค่าพารามิเตอร์ความยาก 3 ประเภท, รูปแบบการทำหน้าที่ต่างกัน มี 4 รูปแบบ คือ Constant DIF, Balanced DIF, Shift low DIF, Shift high DIF และขนาดของการทำหน้าที่ต่างกัน มี 2 ขนาด คือ .1% และ .25%

ผลการศึกษา พบว่า การตรวจสอบด้วยวิธีการ Mantel และวิธีการ GMH เมื่อกำหนดสเกลใหม่และรวมคะแนนของข้อสอบที่ศึกษาไว้ด้วยกัน ให้ผลในการตรวจสอบไม่แตกต่างกัน ในกรณี Constant DIF วิธีการ Mantel มีความถูกต้องในการตรวจสอบสูงกว่าวิธีการ GMH ส่วนกรณี Balanced DIF วิธีการ GMH มีความถูกต้องในการตรวจสอบสูงกว่าวิธีการ Mantel มาก โดยเฉพาะกรณีที่ DIF มีขนาด .25% ตรวจสอบได้ถูกต้องประมาณ 25% ในขณะที่วิธีการ Mantel มีความถูกต้องในการตรวจสอบเพียง 4% ซึ่งสอดคล้องกับสมมติฐานที่ตั้งไว้ เนื่องจากวิธีการ Mantel มีความไวในการตรวจสอบต่อความแตกต่างของค่าเฉลี่ย ซึ่งคาดว่าจะมีน้อยใน DIF กรณีนี้ ในขณะที่วิธีการ GMH มีความไวในการตรวจสอบต่อความแตกต่างระหว่างกลุ่มในทุก ๆ ช่วงของคะแนนสอบ ส่วนกรณี Shift low DIF และ Shift high DIF มีความถูกต้องในการตรวจสอบของทั้งสองวิธี ใกล้เคียงกัน

รูปแบบของ DIF ทุกรูปแบบ ยกเว้นกรณี Constant DIF มีความถูกต้องในการตรวจสอบต่ำมากสำหรับ DIF ที่มีขนาด .1 ส่วน Constant DIF ที่ขนาด .1 มีความถูกต้องในการตรวจสอบ ประมาณ 18% และ 11% สำหรับวิธีการ Mantel และวิธีการ GMH ตามลำดับ ส่วน Constant DIF ขนาด .25 มีความถูกต้องในการตรวจสอบ อยู่ในช่วง 13% ถึง 76% ยกเว้นวิธีการ Mantel ในเงื่อนไข Balanced DIF มีความถูกต้องในการตรวจสอบ เพียง 4% จากการศึกษาของเวลช์และสวเวอร์ ที่ศึกษาจากการจำลองข้อมูลที่เหมือนกันกับการศึกษานี้ พบว่า วิธี Mantel มีความถูกต้องในการตรวจสอบค่อนข้างสูงกว่าผลของการศึกษานี้ ทั้งเงื่อนไขที่เป็นกลางและไม่เป็นกลาง ส่วนเหตุผลที่นำไปได้ที่ทำให้ผลการการศึกษาของเวลช์ และสวเวอร์ ต่างจากการศึกษานี้ครั้งนี้ คือ ข้อสอบที่ศึกษาไม่ได้รวมอยู่กับตัวแปรการจับคู่

เวลช์และมิลเลอร์ (Welch & Miller. 1995 : 163-178) ได้ศึกษาเปรียบเทียบผลของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่า 3 วิธี คือ วิธีการ HW3, วิธีการ GMH และวิธีการ LDFA ภายใต้เงื่อนไขของการรวมกันของตัวแปรจับคู่ภายในและภายนอกที่ต่างกัน 3 เงื่อนไข คือ เงื่อนไข A เป็นการจับคู่โดยใช้คะแนนผลการตอบข้อสอบแบบเลือกตอบ จำนวน 40 ข้อ, เงื่อนไข B เป็นการจับคู่โดยใช้คะแนนผลการตอบข้อสอบแบบเลือกตอบ จำนวน 40 ข้อกับคะแนนจากข้อสอบประเมินการเขียนตอบแบบอธิบายความ, เงื่อนไข C เป็นการจับคู่โดยใช้คะแนนผลการตอบข้อสอบแบบเลือกตอบ จำนวน 40 ข้อกับคะแนนจากข้อสอบประเมินการเขียนตอบทั้งแบบอธิบายความและแบบพรรณนาความ กลุ่มตัวอย่าง

เป็นนักเรียนเกรด 8 แบ่งเป็นกลุ่มย่อย 4 กลุ่ม คือ กลุ่มอเมริกันผิวดำ, กลุ่มอเมริกันผิวขาว, กลุ่มเพศชาย และกลุ่มเพศหญิง เครื่องมือที่ใช้เป็นแบบทดสอบประเมินการเขียนตอบ แบ่งเป็น 3 รูปแบบ คือ 01A, 01B และ 01C โดยรูปแบบ 01A มีผู้สอบจำนวน 1,497 คน, รูปแบบ 01B มีผู้สอบจำนวน 1,523 คน และรูปแบบ 01C มีผู้สอบจำนวน 1,233 คน สัมประสิทธิ์การอ้างอิงของแบบทดสอบการประเมินการเขียนตอบเป็น .60 ส่วนสัมประสิทธิ์ความเชื่อมั่นแบบความสอดคล้องภายใน (แอลฟา) ของข้อสอบแบบเลือกตอบ เป็น .90

ผลการศึกษา พบว่า ผลการประเมินการเขียนจากแบบทดสอบการเขียนและแบบทดสอบเลือกตอบ เพศหญิงทำได้ดีกว่าเพศชาย และอเมริกันผิวขาวทำได้ดีกว่าอเมริกันผิวดำ ซึ่งผลเช่นนี้ยังไม่สามารถสรุปได้ว่ามี DIF เสมอไป ต้องทำการวิเคราะห์ต่อโดยใช้วิธีการที่สามารถแยกผลกระทบ (impact) ออกจากการทำหน้าที่ต่างกันจริง (true DIF) ส่วนค่าสหสัมพันธ์ระหว่างข้อสอบประเมินการเขียนตอบกับข้อสอบแบบเลือกตอบมีความสอดคล้องกันอยู่ในช่วง .66 ถึง .77 ซึ่งบ่งชี้ว่ามีความแตกต่างในการวัดด้วยเครื่องมือวัด 2 ชนิดนี้ แต่เมื่อรวมข้อสอบประเมินการเขียนตอบเข้ากับข้อสอบแบบเลือกตอบ ค่าสหสัมพันธ์ยังมีค่าสูงขึ้น

กรณีที่จำแนกผู้สอบตามเพศ วิธีการทั้ง 3 วิธีสามารถตรวจสอบพบ DIF ได้สอดคล้องกันสูงอย่างมีนัยสำคัญ ในทุกข้อและทุกเงื่อนไขการจับคู่ เมื่อจำแนกตามสีผิวและภายใต้เงื่อนไขการจับคู่แบบ A แล้ว วิธีการตรวจสอบทั้ง 3 วิธีสามารถตรวจสอบพบการทำหน้าที่ต่างกันของข้อสอบ จำนวน 4 ข้อ จากข้อสอบที่มีการทำหน้าที่ต่างกัน จำนวน 6 ข้อ มีข้อสอบ จำนวน 1 ข้อที่ไม่มีวิธีการใดตรวจสอบพบ ภายใต้เงื่อนไข A (ข้อสอบ 01B-1) แต่สามารถตรวจสอบพบ DIF ภายใต้เงื่อนไขอีก 2 เงื่อนไขในทุกวิธีการ มีข้อสอบจำนวน 3 ข้อภายใต้เงื่อนไข A ที่วิธีการ 3 วิธี สามารถตรวจสอบพบ DIF แต่ไม่สอดคล้องกัน และมีบางกรณีภายใต้เงื่อนไข B และ C ของกลุ่มตัวอย่างขนาดเล็กที่เป็นกลุ่มอเมริกันผิวดำ โดยเฉพาะในรูปแบบ 01C ที่ความถูกต้องในการตรวจสอบไม่สอดคล้องกันในระหว่าง 3 วิธีการ และพบว่าในกลุ่มตัวอย่างขนาดเล็ก การแปลผลที่ได้ต้องแปลผลด้วยความระมัดระวัง ผลของการศึกษานี้ แสดงให้เห็นว่า การเพิ่มข้อสอบที่ประเมินการเขียนตอบ เข้าร่วมกับ ข้อสอบแบบเลือกตอบ จะทำให้ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบมีความเชื่อมั่นได้มากขึ้น

ซางและคณะ (Chang et al. 1996 : 333-353) ได้เปรียบเทียบประสิทธิภาพของวิธีการชิปเทสท์ที่ปรับใหม่กับวิธีการ Mantel และวิธีการ Standardized Mean Difference (SMD) โดยใช้การจำลองข้อมูลในการศึกษา 2 ครั้ง การศึกษาครั้งแรกใช้การจำลองข้อมูลจากการสอบ NAEP ซึ่งสวิกค์และคณะ (Zwick et al. 1993) ได้ศึกษาไว้ประกอบด้วยข้อสอบที่ให้คะแนนแบบ 2 ค่า และข้อสอบที่ให้คะแนนแบบหลายค่าเพื่อใช้ในการจับคู่ โดยวิธีการชิปเทสท์ที่ปรับใหม่ใช้ข้อสอบจำนวน 24 ข้อ เป็นข้อสอบที่ให้คะแนนแบบ 2 ค่า จำนวน 20 ข้อ และข้อสอบที่ให้คะแนนแบบหลายค่าแบบ 4 ลำดับชั้นคะแนนคือ 0, 1, 2, 3 จำนวน 4 ข้อ ส่วนวิธีการ Mantel และวิธีการ SMD ใช้ข้อสอบจำนวน 25 ข้อ เป็นข้อสอบที่ให้คะแนนแบบ 2 ค่า จำนวน 20 ข้อ และข้อสอบที่ให้คะแนนแบบหลายค่าแบบ 4 ลำดับชั้นคะแนนคือ 0, 1, 2, 3 จำนวน 5 ข้อ ในการจำลองการตอบข้อสอบใช้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ โดยมีเงื่อนไขที่ศึกษาทั้งหมด 54 เงื่อนไข เป็นการแจกแจงความสามารถของกลุ่มสนใจมี 2 แบบ คือ $N(0,1)$ กับ $N(-1,1)$ ทำการศึกษาทั้งข้อสอบ 27 ข้อ หรือ เท่ากับ $[(4 \times 2 \times 3) + 3]$ ซึ่งเกิดจากประเภทของการทำหน้าที่ต่างกัน (DIF) 4 ประเภท คือ Constant DIF, Low-shift DIF, High-shift DIF และ Balanced DIF, ขนาดของการทำหน้าที่ต่างกัน (DIF) 2 ขนาด คือ .1 และ .25, ค่าพารามิเตอร์ความยาก 3 ค่า และข้อสอบที่เป็น null DIF อีกจำนวน 3 ข้อ แต่ละเงื่อนไขกระทำซ้ำ (replications) จำนวน 600 ครั้ง ผลการศึกษาพบว่า วิธีการชิปเทสท์ที่ปรับใหม่ตรวจสอบการทำหน้าที่ต่างกันได้ดี แต่วิธีการ Mantel และวิธี SMD ตรวจสอบการทำหน้าที่ต่างกันได้ดีกว่าวิธีการ SIBTEST

เล็กน้อย ส่วนความผิดพลาดประเภท 1 ของวิธีการ Mantel และวิธีการ SMD เป็น .049 และ .046 ตามลำดับ ส่วนความผิดพลาดประเภท 1 ของวิธีการ SIBTEST สูงกว่าเล็กน้อย เป็น .063 วิธีการทั้ง 3 วิธีสามารถตรวจสอบการทำหน้าที่ต่างกันแบบสม่ำเสมอได้ดี

การศึกษาครั้งที่สอง เป็นการศึกษาที่เพิ่มเติมจากการศึกษาครั้งแรกให้กว้างขวางขึ้นในเรื่องความหลากหลายของค่าอำนาจจำแนกของข้อสอบที่ศึกษา ผลการศึกษาพบว่า เมื่อพารามิเตอร์อำนาจจำแนกมีค่ามากขึ้น ความถูกต้องในการตรวจสอบสูงขึ้น ปริมาณ DIF ขึ้นอยู่กับพารามิเตอร์อำนาจจำแนกโดยตรง ซึ่งเป็นไปตามที่คาดหวัง เมื่อกำหนดการเปลี่ยนแปลงพารามิเตอร์ความยากไว้คงที่ วิธี SIBTEST สามารถควบคุมผลกระทบที่ก่อให้เกิดความผิดพลาดประเภท 1 ได้เหนือกว่าวิธีการ Mantel และวิธีการ SMD ภายใต้งี้อื่นๆ ใดๆ ไปของการตรวจสอบการทำหน้าที่ต่างกัน เมื่อขนาดของกลุ่มตัวอย่างเพิ่มขึ้น ความถูกต้องในการตรวจสอบของวิธีการ Mantel และวิธีการ SMD จะเพิ่มขึ้นอย่างรวดเร็ว ในขณะที่วิธีการ SIBTEST มีความคงที่พอควร

งานวิจัยในประเทศไทย

กาญจนา วัจนสุนทร (2538) ได้พัฒนาเกณฑ์ในการตัดสินข้อสอบลำเอียงทางเพศ โดยใช้ข้อมูลเชิงประจักษ์จากการสอบคัดเลือกเข้าศึกษาต่อในระดับอุดมศึกษา ปีการศึกษา 2535 โดยใช้วิธีการตรวจสอบ 4 วิธี คือ วิธีตามทฤษฎีการตอบข้อสอบแบบ 2 พารามิเตอร์ โดยแยกเป็นวิธีการหาพื้นที่ระหว่างโค้งของการตอบข้อสอบชนิดคิดเครื่องหมาย (signed area : SA) และชนิดไม่คิดเครื่องหมาย (unsigned area : UA) วิธีแมนเทล-แฮนส์เชล และวิธีซิปเทสต์ ดัชนีที่พัฒนาเพื่อใช้เป็นเกณฑ์ในการตัดสินข้อสอบลำเอียง คือ SA, UA, α_{MH} และ β_{SIB} ตามลำดับ โดยศึกษาภายใต้เงื่อนไขของปัจจัยที่แปรเปลี่ยน 2 ตัว คือ ความยาวของแบบทดสอบ ซึ่งมี 2 วิชา คือ วิชาคณิตศาสตร์ มี 3 ขนาด คือ 20, 30 และ 40 ข้อ ส่วนวิชาภาษาอังกฤษ มี 4 ขนาด คือ 50, 60, 70 และ 80 ข้อ และขนาดของกลุ่มตัวอย่าง มี 6 ขนาด คือ 100, 200, 400, 600, 800 และ 1,000 คน

ผลการวิจัยพบว่า ขนาดของกลุ่มตัวอย่างมีอิทธิพลต่อค่าเฉลี่ยของดัชนีทุกตัว ส่วนความยาวของแบบทดสอบมีอิทธิพลต่อค่าเฉลี่ยของดัชนี SA และ UA แต่ไม่มีอิทธิพลต่อค่าเฉลี่ยของ α_{MH} และ β_{SIB} เกณฑ์ที่พัฒนาขึ้นเพื่อใช้ตัดสินความลำเอียงระหว่างเพศหญิงและเพศชาย มีดังนี้

1. SA > .80 และ UA > .50 เมื่อความยาวของแบบทดสอบน้อยกว่า 50 ข้อ
2. SA > .40 และ UA > 1.20 เมื่อความยาวของแบบทดสอบมากกว่า 50 ข้อขึ้นไป
3. $\alpha_{MH} < .60$, $\alpha_{MH} > 1.40$ และ $\beta_{SIB} > .06$ สำหรับทุกขนาดของกลุ่มตัวอย่างและความยาวของแบบทดสอบ

เรวดี อินทะสระ (2539) ได้ศึกษาความเที่ยงตรงเชิงพยากรณ์ของแบบทดสอบคัดเลือกที่วิเคราะห์ความลำเอียงต่อเพศ ด้วยวิธีทฤษฎีการตอบข้อสอบ (IRT), วิธีแมนเทล-แฮนส์เชล (MH) และวิธีซิปเทสต์ (SIBTEST) พร้อมทั้งศึกษาการตัดสินผลการสอบที่คิดคะแนนมาตรฐานที่ปกติ และ คะแนนน้ำหนักความสามารถและสาเหตุของความลำเอียงของข้อสอบ โดยศึกษาความลำเอียงของข้อสอบคัดเลือกเข้าศึกษาในชั้นปีที่ 1 ประเภทรับตรง ปีการศึกษา 2538 ของมหาวิทยาลัยสงขลานครินทร์ ในวิชาเอกภาษาไทย ก วิชาสังคมศึกษา ก และ วิชาภาษาอังกฤษ กข วิชาละ 8,127 คน (เป็นชาย 2,722 คน หญิง 5,405 คน) วิชาภาษาไทย กข วิชาสังคมศึกษา กข และวิชาภาษาอังกฤษ กข วิชาละ 5,415 คน (เป็นชาย 1,451 คน หญิง 3,961 คน)

ผลการศึกษา พบว่า วิธีการตรวจสอบความลำเอียงทั้ง 3 วิธี ตัดสินจำนวนข้อสอบที่ลำเอียงแตกต่างกันในวิชาภาษาไทย ก ฉบับที่ 2 และวิชาสังคมศึกษา ก ฉบับที่ 1 ที่ระดับนัยสำคัญทางสถิติที่ .05 นอกนั้นแตกต่างกันที่ระดับนัยสำคัญ .01 โดยวิธีทฤษฎีการตอบข้อสอบตัดสินจำนวนข้อสอบที่ลำเอียงได้มากที่สุด ความสัมพันธ์ของลำดับที่ของการสอบ ไม่ว่าจะคิดคะแนนมาตรฐานปกติที่ หรือ คิดคะแนนน้ำหนักความสามารถและใช้ข้อสอบทั้งหมด หรือใช้เฉพาะข้อสอบที่ปราศจากความลำเอียง ต่างมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

เสรี ชัดรัมย์ (2539) ได้เปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอนุกรมระหว่างวิธีแมนเทิล-แฮนส์เซลแบบปกติ กับ วิธีแมนเทิล-แฮนส์เซลแบบแบ่งกลุ่มความสามารถของผู้สอบและความยากของข้อสอบ โดยใช้แบบทดสอบวัดความสามารถในการอ่านภาษาไทย ที่ผู้วิจัยสร้างขึ้นเอง จำนวน 75 ข้อ ศึกษาในกลุ่มตัวอย่างที่เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 จำนวน 1,200 คน จำแนกเป็นเพศชายจำนวน 600 คน เป็นกลุ่มสนใจ และเพศหญิง จำนวน 600 คน เป็นกลุ่มอ้างอิง

ผลการศึกษาพบว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอนุกรม วิธีแมนเทิล-แฮนส์เซลแบบปกติสามารถระบุข้อสอบที่ทำหน้าที่ต่างกันสอดคล้องกับวิธี IRTarea คิดเป็นร้อยละ 33.33 ส่วนวิธีแมนเทิล-แฮนส์เซลแบบแบ่งกลุ่มความสามารถของผู้สอบและความยากของข้อสอบสามารถระบุข้อสอบที่ทำหน้าที่ต่างกันสอดคล้องกับวิธี IRTarea คิดเป็นร้อยละ 61.11

พรรณี จิตมาศ (2540) ได้วิเคราะห์ความลำเอียงต่อเพศของแบบทดสอบคณิตศาสตร์โจทย์ปัญหาที่ผู้วิจัยสร้างขึ้นเอง ด้วยวิธีวิเคราะห์ความลำเอียง 3 วิธี คือ วิธีแปลงค่าความยาก, วิธี MH และวิธี SIBTEST ในแต่ละขนาดของกลุ่มผู้สอบ 500 และ 1000 คนโดยเปรียบเทียบจำนวนข้อที่มีความลำเอียงและเปรียบเทียบค่าความเชื่อมั่นแบบครึ่งฉบับของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 2 ปีการศึกษา 2539 ของโรงเรียนสังกัดกรมสามัญศึกษาส่วนกลางจำนวน 2,200 คน ซึ่งเลือกมาโดยการสุ่มแบบแบ่งชั้น มีขนาดของโรงเรียนเป็นชั้นและโรงเรียนเป็นหน่วยการสุ่ม เครื่องมือที่ใช้คือแบบทดสอบคณิตศาสตร์โจทย์ปัญหาเรื่องสมการ อัตราส่วน และร้อยละ เป็นแบบเลือกตอบชนิด 5 ตัวเลือก จำนวน 40 ข้อ ที่ผู้วิจัยสร้างขึ้นเอง จำนวน 1 ฉบับ

ผลการวิจัยพบว่า เมื่อวิเคราะห์จากกลุ่มตัวอย่างขนาด 500 คน วิธี SIBTEST พบข้อสอบที่มีความลำเอียงมากที่สุดและวิธีแปลงค่าความยากพบข้อสอบที่มีความลำเอียงน้อยที่สุด โดยจำนวนข้อสอบที่มีความลำเอียงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ทุกวิธีวิเคราะห์ และเมื่อวิเคราะห์จากกลุ่มผู้สอบขนาด 1,000 คน วิธี MH พบข้อสอบมีความลำเอียงมากที่สุด วิธีแปลงค่าความยากไม่พบข้อสอบที่ลำเอียง โดยจำนวนข้อสอบที่ลำเอียงจากการวิเคราะห์ด้วยวิธีแปลงค่าความยาก กับ วิธี MH และ วิธีแปลงค่าความยาก กับวิธี SIBTEST แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 นอกนั้นมีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

ญาณภัทร สีหะมงคล (2540) เปรียบเทียบความสอดคล้องของผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันระหว่างวิธี Lord's χ^2 วิธี Raju's Area Measures และ วิธี Closed Interval Area เมื่อขนาดของกลุ่มตัวอย่าง ความยาวของแบบทดสอบ และสัดส่วนจำนวนข้อสอบที่ทำหน้าที่ต่างกันแบบทดสอบต่างกัน ข้อมูลที่ใช้ในการศึกษาเป็นผลการสอบประเมินคุณภาพและความก้าวหน้าทางการศึกษา วิชาคณิตศาสตร์ของนักเรียนชั้นประถมศึกษาปีที่ 4 ปีการศึกษา 2536 ของสำนักงานการประถมศึกษาแห่งชาติ จำนวน 11,404 คน เครื่องมือที่ใช้เป็นแบบทดสอบแบบเลือกตอบ จำนวน 80 ข้อ

ผลการศึกษาพบว่า จำนวนข้อสอบที่ทำหน้าที่ต่างกันจากการตรวจสอบด้วยวิธีการทั้งสาม แตกต่างกัน เมื่อขนาดของกลุ่มตัวอย่างและความยาวของแบบทดสอบต่างกัน ส่วนความสัมพันธ์ระหว่างวิธีการทั้งสาม มีค่าสัมประสิทธิ์สหสัมพันธ์ค่อนข้างสูงมากและมีนัยสำคัญทางสถิติเกือบทุกเงื่อนไขของการศึกษาและสำหรับความสอดคล้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ส่วนมากจะมีค่าปานกลางถึงต่ำ เกือบทุกเงื่อนไขของการศึกษา

อารี วัชรโสติกุล (2543) ได้ศึกษาเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยใช้รูปแบบต่างกัน คือ รูปแบบคะแนนรวมทั้งฉบับ แยกตามเนื้อหา และแยกตามระดับพฤติกรรม ด้วยวิธีการตรวจสอบต่างกัน คือ วิธี SIBTEST และวิธีการถดถอยโลจิสติก แล้วทำการคัดเลือกข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบ เพื่อเปรียบเทียบค่าความเชื่อมั่น

ผลการศึกษาพบว่า จำนวนข้อสอบที่ทำหน้าที่ต่างกันโดยใช้วิธีการตรวจสอบต่างกันแตกต่างกัน ในรูปแบบรวมทั้งฉบับ ส่วนรูปแบบแยกตามเนื้อหาและแยกตามระดับพฤติกรรม ไม่แตกต่างกัน

ทองอยู่ สาระ (2543) ได้ศึกษาเปรียบเทียบอำนาจการตรวจสอบและจำแนกผิดพลาดในการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบสมำเสมอและแบบไม่สมำเสมอ ระหว่างวิธี Mantel-Haenzsel และวิธีการถดถอยโลจิสติก โดยใช้ความยาวแบบทดสอบและขนาดกลุ่มตัวอย่าง ต่างกัน

ผลการศึกษา พบว่า อำนาจการตรวจสอบและการจำแนกผิดพลาดในการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันทั้งแบบสมำเสมอและแบบไม่สมำเสมอระหว่างวิธี Mantel-Haenzsel และวิธีการถดถอยโลจิสติก ภายใต้ความยาวแบบทดสอบและขนาดกลุ่มตัวอย่างที่ศึกษาเกือบทุกเงื่อนไข ไม่แตกต่างกัน ส่วนความยาวของแบบทดสอบ ไม่มีผลต่ออำนาจการตรวจสอบและการจำแนกผิดพลาดในทั้ง 2 วิธี แต่ขนาดของกลุ่มตัวอย่างมีผลต่ออำนาจการตรวจสอบด้วยวิธี Mantel-Haenzsel และวิธีการถดถอยโลจิสติกเกือบทุกเงื่อนไขของการศึกษา คือเมื่อขนาดกลุ่มตัวอย่างเพิ่มขึ้น อำนาจการตรวจสอบจะเพิ่มขึ้น แต่ขนาดของกลุ่มตัวอย่างไม่มีผลต่อการจำแนกผิดพลาดในเกือบทุกเงื่อนไขของการศึกษา

วลีมาศ แซ่อึ้ง (2543) เปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอนเนกรูประหว่างวิธีชิปเทสท์ปรับใหม่ วิธีชิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก ข้อมูลที่ใช้ในการศึกษาจำลองภายใต้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ชนิดกำหนดค่าการเดา (c) คงที่ แล้วจัดกระทำข้อมูลตามปัจจัย 4 ตัว คือ (1) ลักษณะของข้อสอบที่มีค่าความยาก (b) และอำนาจจำแนก (a) ระดับต่ำ ปานกลาง และสูง จำนวน 9 ลักษณะ (2) ความยาวของแบบสอบ 2 ระดับ (3) สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันแบบสอบ 3 ระดับและ (4) ขนาดกลุ่มตัวอย่าง 6 ระดับ รวมข้อมูลที่ศึกษาทั้งหมดจำนวน 324 เงื่อนไข แล้วนำข้อมูลของแต่ละเงื่อนไขมาคำนวณค่าอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอนเนกรูป

ผลการวิจัยพบว่า อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอนเนกรูปของวิธีชิปเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าเท่าเทียมกันภายใต้เกือบทุกเงื่อนไข และทั้งสองวิธีดังกล่าว มีอำนาจการทดสอบสูงกว่าวิธีชิปเทสท์ และวิธีแมนเทล-แฮนส์เซล ภายใต้เกือบทุกเงื่อนไข ส่วนอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอนเนกรูปของวิธีชิปเทสท์ปรับใหม่ วิธีชิปเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก มีค่าอยู่ภายในเกณฑ์ของอัตราความคลาดเคลื่อนประเภทที่ 1 ที่ระดับ 10% ภายใต้เกือบทุกเงื่อนไข

งานวิจัยเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบในระดับบัณฑิตศึกษาในประเทศไทยดังกล่าว เป็น การศึกษาการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบ 2 ค่า (แบบ 0-1) ยังไม่พบว่ามีผู้ใดศึกษา ในประเด็นของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่า · ส่วนตัวแปรที่ ศึกษา ได้แก่ ขนาดของแบบทดสอบ ขนาดของกลุ่มตัวอย่าง อัตราส่วนของกลุ่มประชากรย่อย อัตราส่วนของ ข้อสอบที่ทำหน้าที่ต่างกันในรูปแบบทดสอบ และ วิธีการที่ใช้ในการตรวจสอบ ส่วนการแบ่งกลุ่มประชากรย่อย ส่วนมากจำแนกตามเพศ ภูมิภาค สถานศึกษา เป็นต้น ผู้วิจัยจึงสนใจศึกษาเกี่ยวกับวิธีการ ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่า ระหว่างวิธีการ GMH และวิธีการ Polytomous SIBTEST ในเงื่อนไขขนาดของกลุ่มตัวอย่างและความยาวของแบบทดสอบ

บทที่ 3

วิธีดำเนินการวิจัย

การศึกษาค้นคว้าครั้งนี้มุ่งศึกษาประสิทธิภาพของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในแบบทดสอบที่มีการให้คะแนนแบบหลายค่า เมื่อใช้วิธี GMH และวิธี Polytomous SIBTEST ตรวจสอบว่าจะได้ผลอย่างไร โดยดำเนินการศึกษาตามลำดับ ดังนี้

แหล่งข้อมูล

ข้อมูลที่ใช้ในการศึกษา คือ ข้อมูลผลการตอบข้อสอบวิชาคณิตศาสตร์ เรื่อง วงกลม ของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1 ปีการศึกษา 2545 ของโรงเรียนสังกัดกรมสามัญศึกษา ในจังหวัดพระนครศรีอยุธยา

ประชากร

ประชากรที่ใช้ในการศึกษา เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1 ปีการศึกษา 2545 ของโรงเรียนสังกัดกรมสามัญศึกษา จังหวัดพระนครศรีอยุธยา จำนวน 29 โรงเรียน

กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการศึกษาเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1 ปีการศึกษา 2545 ของโรงเรียนสังกัดกรมสามัญศึกษา จังหวัดพระนครศรีอยุธยา จำนวน 29 โรงเรียน ซึ่งเลือกมาโดยการสุ่มแบบสองขั้นตอน (Two-Stage Random Sampling) มีขนาดของโรงเรียนเป็นชั้น (strata) และมีห้องเรียนเป็นหน่วยของการสุ่ม (Sampling Unit) ซึ่งดำเนินการตามลำดับดังนี้

ขั้นที่ 1 ประมาณขนาดกลุ่มตัวอย่างที่ใช้ในการศึกษาด้วยความเชื่อมั่น 95% ($\alpha = .05$) เมื่อเทียบจากตารางขนาดกลุ่มตัวอย่างที่ควรเลือกจากประชากร พบว่า ต้องใช้กลุ่มตัวอย่าง อย่างน้อย 375 คน (Yamane, 1974 : 403) ในการศึกษาครั้งนี้ต้องการกลุ่มตัวอย่าง เพศชายอย่างน้อย 1,000 คน และกลุ่มตัวอย่างเพศหญิงอย่างน้อย 1,000 คน ใช้เป็นประชากรเทียมในการสุ่มกลุ่มตัวอย่างขนาดต่าง ๆ

ขั้นที่ 2 แบ่งโรงเรียนออกเป็น 3 ขนาดคือ โรงเรียนขนาดใหญ่ โรงเรียนขนาดกลาง และ โรงเรียนขนาดเล็ก ซึ่งข้อมูลโรงเรียนที่ศึกษาจากสำนักงานสามัญศึกษาจังหวัดพระนครศรีอยุธยา พบว่า มีโรงเรียนขนาดใหญ่ 7 โรงเรียน โรงเรียนขนาดกลาง 15 โรงเรียนและโรงเรียนขนาดเล็ก 7 โรงเรียน

ขั้นที่ 3 สุ่มโรงเรียนตามสัดส่วนของโรงเรียนแต่ละขนาด โดยใช้วิธีการสุ่มอย่างง่าย ได้โรงเรียนขนาดใหญ่ 3 โรงเรียน โรงเรียนขนาดกลาง 8 โรงเรียนและโรงเรียนขนาดเล็ก 3 โรงเรียน และสุ่มห้องเรียนจากแต่ละโรงเรียน ด้วยวิธีการสุ่มอย่างง่ายตามสัดส่วนของจำนวนนักเรียน เพื่อให้ได้นักเรียนกลุ่มตัวอย่างตามจำนวนที่ต้องการ รายชื่อโรงเรียน, ห้องเรียน และจำนวนนักเรียน ดังแสดงในตาราง 6

ตาราง 6 รายชื่อโรงเรียน จำนวนห้องเรียน และจำนวนนักเรียน

โรงเรียน	จำนวน ห้องเรียน	จำนวนนักเรียน		
		ชาย	หญิง	รวม
ขนาดใหญ่				
1. ท่าเรือ"นิตยานุกูล"	8	153	167	320
2. บางปะอิน"ราชานุเคราะห์1"	8	134	198	332
3. เสนา"เสนาประสิทธิ์"	8	155	187	342
ขนาดกลาง				
1. นครหลวง"อุดมรัชต์วิทยา"	3	48	54	102
2. ท่าช้างวิทยาคม	3	60	65	125
3. บางไทรวิทยา	4	69	89	158
4. บางปะหัน	6	103	112	215
5. บ้านแพรกประชาสรรค์	3	45	48	93
6. วัดโพธิ์ผักไห่(เวชพันธ์อนุสรณ์)	3	49	56	105
7. วิเชียรกลิ่นสุคนธ์อุปถัมภ์	4	64	79	143
8. วังน้อย(พนมยงค์วิทยา)	3	52	59	111
ขนาดเล็ก				
1. มหาราช"ประชานิมิต"	2	32	38	70
2. ลาดชะโดสามัคคี	2	29	35	64
3. ปากกรานพิทยา	2	32	35	67
รวม	59	1,025	1,222	2,247

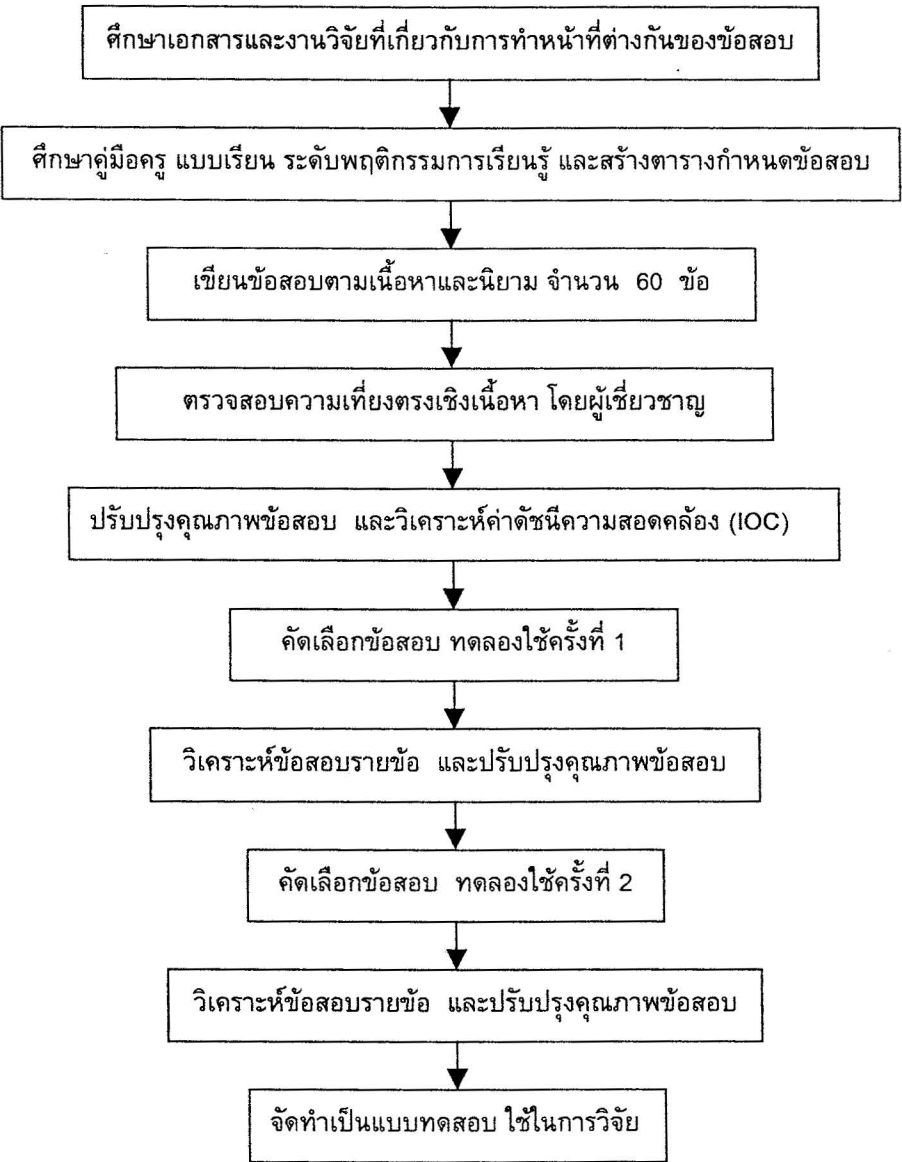
เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 เรื่อง วงกลม ที่ผู้วิจัยสร้างขึ้นมีลักษณะเป็นข้อสอบปรนัยแบบเลือกตอบแบบถูกผิด (Multiple True-False Item) ชนิด 5 ตัวเลือก โดยให้คะแนนทุกตัวเลือก จำนวน 40 ข้อ มีข้อสอบแสดงการทำหน้าที่ต่างกันเอื้อต่อเพศชายจำนวน 8 ข้อ

สำหรับแบบทดสอบขนาด 30 ข้อและ 20 ข้อ ที่ใช้ในการศึกษา ใช้ข้อมูลผลการตอบข้อสอบจากแบบทดสอบที่ประกอบด้วยข้อสอบ 40 ข้อของนักเรียนกลุ่มตัวอย่างที่กำหนดเป็นประชากรเทียบมมาคัดเลือกข้อสอบ โดยการสุ่มอย่างง่าย เป็นข้อสอบที่แสดงการทำหน้าที่ต่างกันที่เอื้อต่อเพศชาย จำนวน 4 ข้อและ 2 ข้อ ตามลำดับ โดยการสุ่มตามสัดส่วน

การสร้างเครื่องมือในการวิจัย

ในการสร้างเครื่องมือสำหรับการวิจัยครั้งนี้ ผู้วิจัยได้ดำเนินการสร้างตามขั้นตอนดังภาพประกอบ 2



ภาพประกอบ 2 ขั้นตอนการสร้างเครื่องมือที่ใช้ในการวิจัย

จากแผนภูมิการดำเนินการสร้างเครื่องมือที่ใช้ในการวิจัย มีรายละเอียดของการดำเนินการสร้างดังนี้

1. ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบ เพื่อศึกษาลักษณะของข้อสอบที่ทำหน้าที่ต่างกัน ซึ่งพอสรุปได้ดังนี้

ลักษณะของข้อสอบโดยทั่วไปที่แสดงการทำหน้าที่ต่างกัน (สิริรัตน์ วิภาสศิลป์, 2545 : 64)

1. เนื้อหาหรือภาษาในข้อสอบยั่วให้ผู้ตอบข้อสอบกลุ่มสนใจโกรธ เกิดการโต้แย้งหรือเกิดอารมณ์ไม่พอใจ
2. เนื้อหาหรือภาษาในข้อสอบมีความหมายไปในทางลบ ดูถูกเหยียดหยามหรือก้าวร้าวต่อผู้ตอบข้อสอบกลุ่มสนใจ
3. เนื้อหาหรือภาษาในข้อสอบแสดงว่าผู้ตอบข้อสอบกลุ่มสนใจมีปมด้อยเกี่ยวกับอำนาจหรือความเป็นผู้นำ
4. เนื้อหาหรือภาษาในข้อสอบหลาย ๆ ข้อให้ความสนใจ เน้นความสำคัญและยกย่องผู้ตอบกลุ่มอ้างอิง
5. เนื้อหาในข้อสอบมีสารสนเทศที่เป็นประโยชน์กับกลุ่มอ้างอิงมากกว่ากลุ่มสนใจ

ลักษณะข้อสอบที่แสดงการทำหน้าที่ต่างกันทางเพศ

1. รูปแบบหรือโครงสร้างของข้อสอบเป็นปัญหากับผู้ตอบข้อสอบเพศใดเพศหนึ่งมากกว่าผู้ตอบข้อสอบอีกเพศหนึ่ง
2. เนื้อหาในข้อสอบใช้สรรพนาม เป็นการให้ความสนใจความสำคัญและยกย่องเฉพาะเพศใดเพศหนึ่ง
3. เนื้อหาในข้อสอบกำหนดสถานการณ์ที่เพศใดเพศหนึ่งสนใจและได้รับการปลูกฝัง ผักผ่อน หรือ มีโอกาสพบเห็นในชีวิตประจำวันมากกว่า

ส่วนลักษณะของข้อสอบคณิตศาสตร์ที่แสดงการทำหน้าที่ต่างกันทางเพศ (Gipps & Murphy, 1994 : 182) คือ

1. ลักษณะของโจทย์ที่มีความซับซ้อนและท้าทาย เพศชายจะชอบมากกว่าเพศหญิง
2. ลักษณะของโจทย์ที่เป็นโจทย์ปัญหาหรือการหาความสัมพันธ์และความเข้าใจเกี่ยวกับขนาดรูปทรงและมิติต่าง ๆ การคาดคะเนระยะทาง การมองรูปทรง (มิติสัมพันธ์) เพศชายจะชอบมากกว่าเพศหญิง
3. ลักษณะของโจทย์ที่ยากต้องใช้ความพยายามในการค้นหาคำตอบ เพศชายจะชอบมากกว่าเพศหญิง
4. ลักษณะของตัวเลือกที่ต้องวิเคราะห์ซับซ้อน เพศชายจะชอบมากกว่าเพศหญิง ส่วนเพศหญิงชอบตัวเลือกที่ระบุว่า “ไม่มีคำตอบถูก” หรือ “ถูกทุกข้อ” มากกว่าเพศชาย

จากการศึกษาของ สมิธและวอคเกอร์ (Garmer & Engelhard, 1999 : 30 citing Smith & Walker, 1988) พบว่า เพศหญิงจะได้เปรียบในการทำข้อสอบพีชคณิต ส่วนเพศชายจะได้เปรียบในการทำข้อสอบเรขาคณิต และการศึกษาของเบอร์ตัน (Burton, 1996); กัลแลคเกอร์ (Gallagher, 1992); โอเนลและแมคพีค (O'Neill & McPeck, 1993) และดูลิตเติลและเคลียร์ (Doolittle & Cleary, 1987) (Garmer & Engelhard, 1999 : 31 citing Burton, 1996; Gallagher, 1992; O'Neill & McPeck, 1993 and Doolittle & Cleary, 1987) พบว่า ลักษณะของข้อสอบคณิตศาสตร์ที่เพศชายจะทำได้ดีกว่าเพศหญิงได้แก่ ข้อสอบเรขาคณิตและพีชคณิตที่มีเหตุผลเชิงคณิตศาสตร์, งานที่มีลักษณะเป็นการแก้ปัญหา (Problem Solving) และ งานที่ใช้ความคิดในขั้นสูง (higher level cognitive), ข้อสอบเกี่ยวกับอัตราส่วน (ratio), สัดส่วน (proportions) หรือ เปอร์เซนต์ (percent) และข้อสอบที่มีตาราง, กราฟ หรือ ภาพ ส่วนลักษณะของข้อสอบคณิตศาสตร์ที่เพศหญิงจะทำได้ดีกว่าเพศชาย ได้แก่ การคิดคำนวณ, พีชคณิต, ข้อสอบที่ใช้สัญลักษณ์และข้อสอบคณิตศาสตร์ที่เป็นนามธรรม (abstract mathematics) และงานที่มีลักษณะคล้ายคลึงกับในตำรา

จากลักษณะดังกล่าวข้างต้น ผู้วิจัยจึงกำหนดเงื่อนไขในการสร้างแบบทดสอบวิชาคณิตศาสตร์ เรื่อง วงกลม ที่แสดงการทำหน้าที่ต่างกันเพื่อต่อเพศชาย ดังนี้

1) ลักษณะของโจทย์ปัญหาที่ซับซ้อน ต้องพิจารณาโจทย์ปัญหาแล้วแปลงให้เป็นภาพ ต้องใช้ความสามารถทางมิติสัมพันธ์ในการมองรูปทรง การคาดคะเนความยาว ระยะทาง การเชื่อมโยงสิ่งที่กำหนด และต้องใช้ความพยายามในการค้นหาคำตอบ

2) ลักษณะของตัวเลือกที่ต้องวิเคราะห์ซับซ้อนในการค้นหาคำตอบ

2. ศึกษาคู่มือครูและแบบเรียนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 เรื่อง วงกลม เพื่อพิจารณาโครงสร้างหลักสูตร กำหนดเนื้อหา และจำนวนข้อสอบ มีรายละเอียดดังนี้

วิชาคณิตศาสตร์ เรื่อง วงกลม ซึ่งเป็นเนื้อหาเชิงเรขาคณิต มีเนื้อหาสาระย่อย ๆ พอสรุปเป็นประเภท ได้ดังนี้

1. ประเภทศัพท์และนิยาม ได้แก่ วงกลม เส้นผ่านศูนย์กลาง ครึ่งวงกลม วงกลมที่เท่ากัน คอร์ด เส้นสัมผัส มุมในครึ่งวงกลม มุมในส่วนโค้งของวงกลม มุมที่จุดศูนย์กลาง และรูปหลายเหลี่ยม

2. ประเภทหลักการ ทฤษฎี หรือ กฎ ได้แก่ สมบัติของวงกลมต่าง ๆ และสมบัติของรูปหลายเหลี่ยมด้านเท่ามุมเท่า ต่าง ๆ

การวิเคราะห์ระดับพฤติกรรมการเรียนรู้ด้านความรู้และการคิด วิชาคณิตศาสตร์ เรื่อง วงกลม นี้ ยึดตามแนวคิดของวิลสัน (Wilson, 1971 : 645-694) ซึ่งแบ่งเป็น 4 ระดับ คือ

1. ความรู้ ความจำด้านการคิดคำนวณ (Computation) แบ่งเป็น 3 ชั้น คือ

1.1 ความรู้ ความจำเกี่ยวกับข้อเท็จจริง (Specific facts)

1.2 ความรู้ ความจำเกี่ยวกับศัพท์และนิยาม (Terminology)

1.3 ทักษะในการใช้กระบวนการคิดคำนวณ (Algorithms)

2. ความเข้าใจ (Comprehension) แบ่งเป็น 6 ชั้น คือ

2.1 ความเข้าใจเกี่ยวกับมโนภาพ (Concept)

2.2 ความเข้าใจเกี่ยวกับหลักการ กฎทางคณิตศาสตร์ และการทำให้เป็นรูปทั่วไป
(Principles rules and generalization)

2.3 ความเข้าใจเกี่ยวกับโครงสร้างทางคณิตศาสตร์ (Mathematics structure)

2.4 ความเข้าใจเกี่ยวกับการแปลงโจทย์ปัญหา (Transform problem elements)

2.5 ความเข้าใจในการติดตามแนวของเหตุผล (Follow a line of reasoning)

3. การนำไปใช้ (Application) แบ่งเป็น 4 ชั้น คือ

3.1 ความสามารถในการแก้ปัญหาธรรมดา (Routine problem)

3.2 ความสามารถในการเปรียบเทียบ (Comparison)

3.3 ความสามารถในการวิเคราะห์ข้อมูล (Analyze data)

3.4 ความสามารถในการมองเห็นแบบ โครงสร้างที่เหมือนกันและสมมาตร (Recognize patterns, isomorphism's and symmetries)

4. การวิเคราะห์ (Analysis) แบ่งเป็น 5 ชั้น คือ

4.1 ความสามารถในการแก้ปัญหาโจทย์พลิกแพลง (Non-routine problem)

4.2 ความสามารถในการค้นพบความสัมพันธ์ (Discover relationship)

- 4.3 ความสามารถในการพิสูจน์ (Construct proof)
- 4.4 ความสามารถในการวิจารณ์การพิสูจน์ (Criticize proof)
- 4.5 ความสามารถเกี่ยวกับการสร้างสูตรและตรวจสอบความถูกต้องของสูตร (Formulate and validate generalization)

ด้วยจุดมุ่งหมายของการวิจัยที่ต้องการเปรียบเทียบวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่า โดยใช้แบบทดสอบเป็นเครื่องมือเพียงเพื่อให้ได้ข้อมูลผลการตอบข้อสอบที่มีการให้คะแนนแบบหลายค่า โดยใช้เนื้อหาสาระย่อยของวิชาคณิตศาสตร์ เรื่อง วงกลม และแนวคิดของวิลสัน (Wilson) ในการวิเคราะห์ระดับพฤติกรรมการเรียนรู้ด้านความรู้และการคิด มาเป็นองค์ประกอบสำคัญในการสร้างแบบทดสอบ

จากเนื้อหาสาระย่อยๆ และแนวคิดดังกล่าวข้างต้นและลักษณะของแบบทดสอบที่ต้องการสร้าง คือ เป็นข้อสอบปรนัยแบบเลือกตอบแบบถูกผิด ชนิด 5 ตัวเลือก ข้อสอบแต่ละข้อกำหนดให้ผู้สอบต้องพิจารณาและตอบทุกตัวเลือก ซึ่งในแต่ละตัวเลือกสามารถวัดความรู้ในเนื้อหาสาระย่อยๆ ได้หลากหลาย นั่นคือ ข้อสอบข้อหนึ่งๆ สามารถวัดความรู้ในเนื้อหาสาระได้หลายเนื้อหาและสามารถวัดระดับพฤติกรรมการเรียนรู้ได้หลายระดับ ดังนั้น การวิเคราะห์ว่าข้อสอบข้อนี้วัดระดับพฤติกรรมการเรียนรู้ใด พิจารณาจากการวัดพฤติกรรมของแต่ละตัวเลือก ระดับพฤติกรรมการเรียนรู้ที่สูงสุดของตัวเลือกในข้อนั้น ถือเป็นระดับพฤติกรรมการเรียนรู้ที่ข้อสอบข้อนี้วัดได้ และการวิเคราะห์ว่าข้อสอบข้อนี้วัดเนื้อหาใด พิจารณาโดยภาพรวมว่าเป็นการวัดความรู้เกี่ยวกับวงกลม สมบัติของวงกลมและการแก้โจทย์ปัญหา รายละเอียดแสดงในตาราง 7

ตาราง 7 ตารางกำหนดข้อสอบวิชาคณิตศาสตร์ เรื่องวงกลม จำแนกตามระดับพฤติกรรมที่วัด

เนื้อหา	พฤติกรรมที่วัด (ข้อ)				รวมจำนวนข้อสอบ (ข้อ)
	ความรู้ความจำ	ความเข้าใจ	การนำไปใช้	การวิเคราะห์	
ความรู้เกี่ยวกับวงกลม , สมบัติของวงกลม และการแก้โจทย์ปัญหา	2 (2) 2	19 (14) 11	15 (10) 4	4 (4) 3	40 (30) 20

หมายเหตุ ตัวเลขในวงเล็บ () หมายถึง จำนวนข้อสอบในแบบทดสอบขนาด 30 ข้อ

ตัวเลขที่ขีดเส้นใต้ หมายถึง จำนวนข้อสอบในแบบทดสอบขนาด 20 ข้อ

3. เขียนข้อสอบตามตารางที่กำหนดไว้ใน ข้อ 2 โดยมีจำนวนข้อสอบมากกว่าที่ต้องการใช้จริงประมาณ 20 ข้อ จัดพิมพ์ข้อสอบที่สร้างแล้วนำไปให้ผู้เชี่ยวชาญซึ่งมีความรู้ในเนื้อหาและการเขียนข้อสอบวิชาคณิตศาสตร์และมีความรู้ทางการวัดผลการศึกษามาแล้วไม่ต่ำกว่า 5 ปี จำนวน 5 ท่าน (รายชื่อแสดงในภาคผนวก ก) เป็นผู้ตรวจสอบความเที่ยงตรงเชิงเนื้อหาของแบบทดสอบเป็นรายข้อและการแสดงการทำหน้าที่ต่างกันของข้อสอบ (รายละเอียดในภาคผนวก จ)

4. ปรับปรุงแก้ไขข้อสอบตามข้อเสนอแนะของผู้เชี่ยวชาญ และ วิเคราะห์ดัชนีความสอดคล้องของความคิดเห็นของผู้เชี่ยวชาญเกี่ยวกับความสอดคล้องระหว่างเนื้อหาข้อสอบแต่ละข้อกับระดับพฤติกรรมและ

เกี่ยวกับการแสดงการทำหน้าที่ต่างกันของข้อสอบ โดยใช้ดัชนีความสอดคล้องของโรวินेलลีและแฮมเบิลตัน (บุญเชิด ภิญญอนันตพงษ์. 2527 : 67-70 ; อ้างอิงจาก Rovinelli and Hambleton. 1972) ซึ่งมีสูตรดังนี้

$$IOC = \frac{\sum R}{N}$$

เมื่อ IOC แทน ดัชนีความสอดคล้องของความคิดเห็นของผู้เชี่ยวชาญ
 $\sum R$ แทน ผลรวมของการตัดสินของผู้เชี่ยวชาญ
 N แทน จำนวนผู้เชี่ยวชาญ

โดยผู้เชี่ยวชาญพิจารณาว่า ข้อสอบวัดได้ ตรง กับระดับพฤติกรรมที่กำหนด ตามเกณฑ์ ดังนี้

- 1 แทน แน่ใจว่า ข้อสอบวัดได้ ตรง กับระดับพฤติกรรมที่กำหนด
- 0 แทน ไม่แน่ใจว่า ข้อสอบวัดได้ ตรง กับระดับพฤติกรรมที่กำหนด
- 1 แทน แน่ใจว่า ข้อสอบวัดได้ ไม่ตรง กับระดับพฤติกรรมที่กำหนด

และ มีเกณฑ์ในการพิจารณาการทำหน้าที่ต่างกันของข้อสอบ ดังนี้

- 1 แทน แน่ใจว่า ข้อสอบแสดงการทำหน้าที่ต่างกัน
- 0 แทน ไม่แน่ใจว่า ข้อสอบแสดงการทำหน้าที่ต่างกัน
- 1 แทน แน่ใจว่า ข้อสอบไม่แสดงการทำหน้าที่ต่างกัน

การแปลความหมายดัชนีความสอดคล้องที่คำนวณได้โดยถือว่า ถ้าค่า IOC มีค่าตั้งแต่ 0.5 ขึ้นไป แสดงว่า ข้อสอบวัดได้ตรงกับระดับพฤติกรรมที่กำหนดและข้อสอบแสดงการทำหน้าที่ต่างกัน ซึ่งในการศึกษาครั้งนี้ ได้ค่าดัชนีความสอดคล้องของความคิดเห็นของผู้เชี่ยวชาญ เกี่ยวกับความสอดคล้องระหว่างเนื้อหาข้อสอบแต่ละข้อกับระดับพฤติกรรม ตั้งแต่ 0.6 ถึง 1.0 และได้ค่าดัชนีความสอดคล้องของความคิดเห็นของผู้เชี่ยวชาญเกี่ยวกับการแสดงการทำหน้าที่ต่างกันของข้อสอบเท่ากับ 1 ทุกข้อ (รายละเอียดในภาคผนวก จ)

5. คัดเลือกข้อสอบ จำนวน 50 ข้อ ที่มีค่าดัชนีความสอดคล้องของความคิดเห็นของผู้เชี่ยวชาญ (IOC) ตั้งแต่ 0.5 ขึ้นไป และสอดคล้องกับตารางกำหนดข้อสอบ ซึ่งข้อสอบจำนวน 50 ข้อนี้ประกอบด้วยข้อสอบที่แสดงการทำหน้าที่ต่างกันตามเกณฑ์การพิจารณาในข้อ 4 จำนวน 13 ข้อ จัดพิมพ์และนำไปทดลองใช้ทดสอบครั้งที่ 1 กับ นักเรียนชั้นมัธยมศึกษาปีที่ 3 โรงเรียนอยุธยาวิทยาลัย จังหวัดพระนครศรีอยุธยา จำนวน 2 ห้องเรียน จำนวน 77 คน และโรงเรียนจอมสุรางค์อุปถัมภ์ จังหวัดพระนครศรีอยุธยา จำนวน 1 ห้องเรียน จำนวน 40 คน รวมเป็น 117 คน นำผลการสอบมาวิเคราะห์คุณภาพข้อสอบรายข้อและค่าความเชื่อมั่นของแบบทดสอบ มีค่าความยาก ตั้งแต่ 0.49 ถึง 0.91 และมีค่าอำนาจจำแนก ตั้งแต่ 0.10 ถึง 0.57 และค่าความเชื่อมั่นของแบบทดสอบ เท่ากับ 0.8839 แล้วคัดเลือกข้อสอบโดยพิจารณาค่าสถิติจากผลการวิเคราะห์รายข้อที่มีค่าความยากระหว่าง .20 กับ .80 และค่าอำนาจจำแนกตั้งแต่ .20 ขึ้นไป มาปรับปรุงความชัดเจนของข้อคำถาม เงื่อนไขที่กำหนด และตัวเลือกให้เหมาะสมยิ่งขึ้น

6. คัดเลือกข้อสอบที่ผ่านการพิจารณาค่าสถิติจากข้อ 5 นำ และผ่านการปรับปรุงแล้ว จำนวน 40 ข้อ ซึ่งประกอบด้วยข้อสอบที่แสดงการทำหน้าที่ต่างกันตามเกณฑ์การพิจารณาในข้อ 4 จำนวน 8 ข้อ (รายละเอียดในภาคผนวก จ) จัดพิมพ์และนำไปทดลองใช้ทดสอบครั้งที่ 2 กับนักเรียนชั้นมัธยมศึกษาปีที่ 3 โรงเรียนภาษี “สุนทรวิทยานุกูล” โรงเรียนผักไห้ “สุทธาประมุข” โรงเรียนบางปะอิน และโรงเรียนบางช้าง จังหวัดพระนครศรีอยุธยา โรงเรียนละ 1 ห้องเรียน รวมจำนวนนักเรียน 214 คน นำผลการสอบมาวิเคราะห์

หาคูณภาพของข้อสอบรายข้อ และค่าความเชื่อมั่นของแบบทดสอบ มีค่าความยาก ตั้งแต่ 0.50 ถึง 0.80 และมีค่าอำนาจจำแนก ตั้งแต่ 0.20 ถึง 0.61 และค่าความเชื่อมั่นของแบบทดสอบ เท่ากับ 0.8846 ตรวจสอบค่าสถิติของข้อสอบรายข้อ และ ปรับปรุงข้อสอบให้เหมาะสมยิ่งขึ้น

7. นำข้อสอบที่ผ่านการวิเคราะห์และการปรับปรุงแล้วจากข้อ 6 จำนวน 40 ข้อ จัดพิมพ์เป็นแบบทดสอบฉบับจริง (รายละเอียดในภาคผนวก ง) เพื่อนำไปใช้เป็นเครื่องมือในการเก็บรวบรวมข้อมูล

ลักษณะของเครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 เรื่อง วงกลม มีลักษณะเป็นข้อสอบปรนัยแบบเลือกตอบแบบถูกผิด (Multiple True-False Item) ชนิด 5 ตัวเลือก โดยให้คะแนนทุกตัวเลือก จำนวน 40 ข้อ ที่ผู้วิจัยสร้างขึ้น ซึ่งมีลักษณะดังนี้

ตัวอย่างข้อสอบ

ข้อ (A) กำหนดให้ O เป็นจุดศูนย์กลางของวงกลม A, B, C และ D เป็นจุดบนเส้นรอบวง

และ $\overline{AC}$ เป็นเส้นผ่านศูนย์กลาง ถ้า $\widehat{ADO} = 20^\circ$ แล้ว

ก. $\widehat{DAO} = 20^\circ$

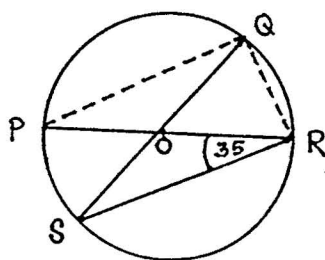
ข. $\widehat{ODC} = 70^\circ$

ค. $\widehat{ABC} = 90^\circ$

ง. $\widehat{DOC} = 140^\circ$

จ. $\widehat{AOD} = 2(\widehat{ACD})$

ข้อ (B) กำหนดให้ O เป็นจุดศูนย์กลางของวงกลม และ $\widehat{PRS} = 35^\circ$



ก. $\widehat{PQO} = 35^\circ$

ข. $\widehat{QOR} = 65^\circ$

ค. $\widehat{OQR} = 35^\circ$

ง. $\widehat{QPO} = \widehat{OSR}$

จ. คอร์ด PQ = คอร์ด RS

ตัวอย่างกระดาษคำตอบ

ข้อ	ก	ข	ค	ง	จ
A	✓	✓	✓	X	✓
B	✓	X	X	✓	✓

จากตัวอย่างข้อสอบ ข้างต้นนั้น ลักษณะของข้อสอบข้อ B เป็นข้อสอบที่โจทย์กำหนดเงื่อนไขให้ทั้งลักษณะที่เป็นข้อความและที่เป็นภาพประกอบ ซึ่งผู้สอบไม่ต้องใช้ความสามารถในการแปลงข้อกำหนดที่เป็นข้อความให้เป็นภาพอีกครั้ง ทำให้ง่ายในการพิจารณาข้อคำถาม สามารถพิจารณาหาคำตอบได้ทันที ส่วน

ลักษณะของโจทย์คำถามของข้อสอบข้อ A นั้น โจทย์กำหนดเงื่อนไขให้เพียงข้อความ ผู้สอบต้องใช้ความสามารถทางมิติสัมพันธ์ในการมองภาพ รูปทรง คาคะเนความยาว ระยะทาง และ การเชื่อมโยงสิ่งที่กำหนดเพื่อแปลงข้อกำหนดที่เป็นข้อความให้เป็นรูปภาพ ซึ่งในการวิจัยครั้งนี้ถือว่าเป็นข้อสอบที่แสดงการทำหน้าที่ต่างกันที่เอื้อต่อเพศชาย

การตรวจให้คะแนน

การตรวจให้คะแนนของแบบทดสอบวิชาคณิตศาสตร์ เรื่อง วงกลม ที่มีลักษณะเป็นข้อสอบปรนัยแบบเลือกตอบแบบถูกผิด (Multiple True-False Item) ชนิด 5 ตัวเลือก ซึ่งผู้ตอบต้องพิจารณาตอบทุกตัวเลือก ถ้าการตอบตรงกับเฉลยทุกตัวเลือก ให้ 5 คะแนน ถ้าการตอบตรงกับเฉลยเพียง 4 ตัวเลือก ให้ 4 คะแนน ถ้าการตอบตรงกับเฉลยเพียง 3 ตัวเลือก ให้ 3 คะแนน ถ้าการตอบตรงกับเฉลยเพียง 2 ตัวเลือก ให้ 2 คะแนน ถ้าการตอบตรงกับเฉลยเพียง 1 ตัวเลือก ให้ 1 คะแนน และถ้าการตอบไม่ตรงกับเฉลยทุกตัวเลือก ให้ 0 คะแนน

การเก็บรวบรวมข้อมูล

ผู้วิจัยดำเนินการ ดังนี้

1. ขอจดหมายจากบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ถึง ผู้อำนวยการสำนักงานสามัญศึกษา เพื่อขอความร่วมมือในการเก็บรวบรวมข้อมูลจากโรงเรียนที่เป็นกลุ่มตัวอย่าง ซึ่งเป็นโรงเรียนในสังกัดกรมสามัญศึกษา จังหวัดพระนครศรีอยุธยา

2. นำหนังสือขอความร่วมมือในการเก็บรวบรวมข้อมูลจากผู้อำนวยการสำนักงานสามัญศึกษา ไปติดต่อประสานงานกับโรงเรียนที่เป็นกลุ่มตัวอย่าง โดย

2.1 ติดต่อกับผู้อำนวยการ หรือผู้ช่วยผู้อำนวยการฝ่ายวิชาการ หรือหัวหน้าหมวดวิชา หรือครูผู้สอนวิชาคณิตศาสตร์ ระดับชั้น มัธยมศึกษาปีที่ 3 เพื่อขอความอนุเคราะห์ในการเก็บรวบรวมข้อมูลในด้านการนัดหมายเวลาและสถานที่ที่ใช้ในการดำเนินการสอบ

2.2 ติดต่อกับครูผู้สอนวิชาคณิตศาสตร์ ระดับชั้น มัธยมศึกษาปีที่ 3 เพื่อชี้แจงจุดประสงค์ของการศึกษา ขอบเขตเนื้อหาที่ต้องการวัด รูปแบบการวัดหรือลักษณะของข้อสอบที่ใช้วัด รวมทั้งจุดเด่น และประโยชน์ของข้อสอบที่สามารถนำไปใช้ในการพัฒนาการเรียนการสอน และเพื่อขอความร่วมมือในการดำเนินการสอนและการนัดหมายการสอบ ซึ่งครูผู้สอนทุกท่านให้ความสนใจกับลักษณะของข้อสอบมาก และ ทุกโรงเรียนให้ความร่วมมือในการเก็บรวบรวมข้อมูลเป็นอย่างดี

ทั้งนี้ ผู้วิจัยเป็นผู้ไปเก็บรวบรวมข้อมูลด้วยตนเอง ทั้งกรณีการทดลองใช้แบบทดสอบและกรณีเก็บรวบรวมข้อมูลจริง

การเตรียมข้อมูล

เมื่อเก็บรวบรวมข้อมูลจากกลุ่มตัวอย่างแล้ว เพื่อให้การวิเคราะห์ข้อมูลเป็นไปตามจุดมุ่งหมายการศึกษา ผู้วิจัยได้ดำเนินการจัดเตรียมข้อมูล ตามขั้นตอนดังนี้

1. คัดเลือกกระดาษคำตอบของนักเรียนที่ตอบครบทุกข้อทุกตัวเลือก จากแบบทดสอบวิชาคณิตศาสตร์ ที่ประกอบด้วยข้อสอบ 40 ข้อ ได้กระดาษคำตอบที่เป็นผลการตอบข้อสอบของนักเรียนกลุ่มตัวอย่างที่กำหนดให้เป็นประชากรเทียมจำนวน 2,247 คน เป็นเพศชาย 1,025 คน เพศหญิง 1,222 คน ตรวจให้

คะแนน โดยตรวจให้คะแนนทุกตัวเลือก ตัวเลือกใดที่ตอบถูกให้เป็น 1 คะแนน ตัวเลือกใดตอบผิดให้เป็น 0 คะแนน คะแนนของข้อสอบแต่ละข้อเป็น 0 ถึง 5 คะแนน คือ ถ้าตอบผิดทุกตัวเลือกได้ 0 คะแนน ถ้าตอบถูก 1 ตัวเลือกได้ 1 คะแนน ถ้าตอบถูก 2 ตัวเลือกได้ 2 คะแนน ถ้าตอบถูก 3 ตัวเลือกได้ 3 คะแนน ถ้าตอบถูก 4 ตัวเลือกได้ 4 คะแนน และถ้าตอบถูกทุกตัวเลือกได้ 5 คะแนน แล้วบันทึกแฟ้มข้อมูลคะแนนผลการตอบข้อสอบของแบบทดสอบฉบับที่มีความยาว 40 ข้อ โดยใช้โปรแกรม SPSS for Windows รวมทั้งบันทึกแฟ้มข้อมูลคะแนนผลการตอบข้อสอบของแบบทดสอบที่มีความยาว 30 ข้อ และ 20 ข้อ ตามลำดับ

2. การเตรียมข้อมูลของกลุ่มตัวอย่าง

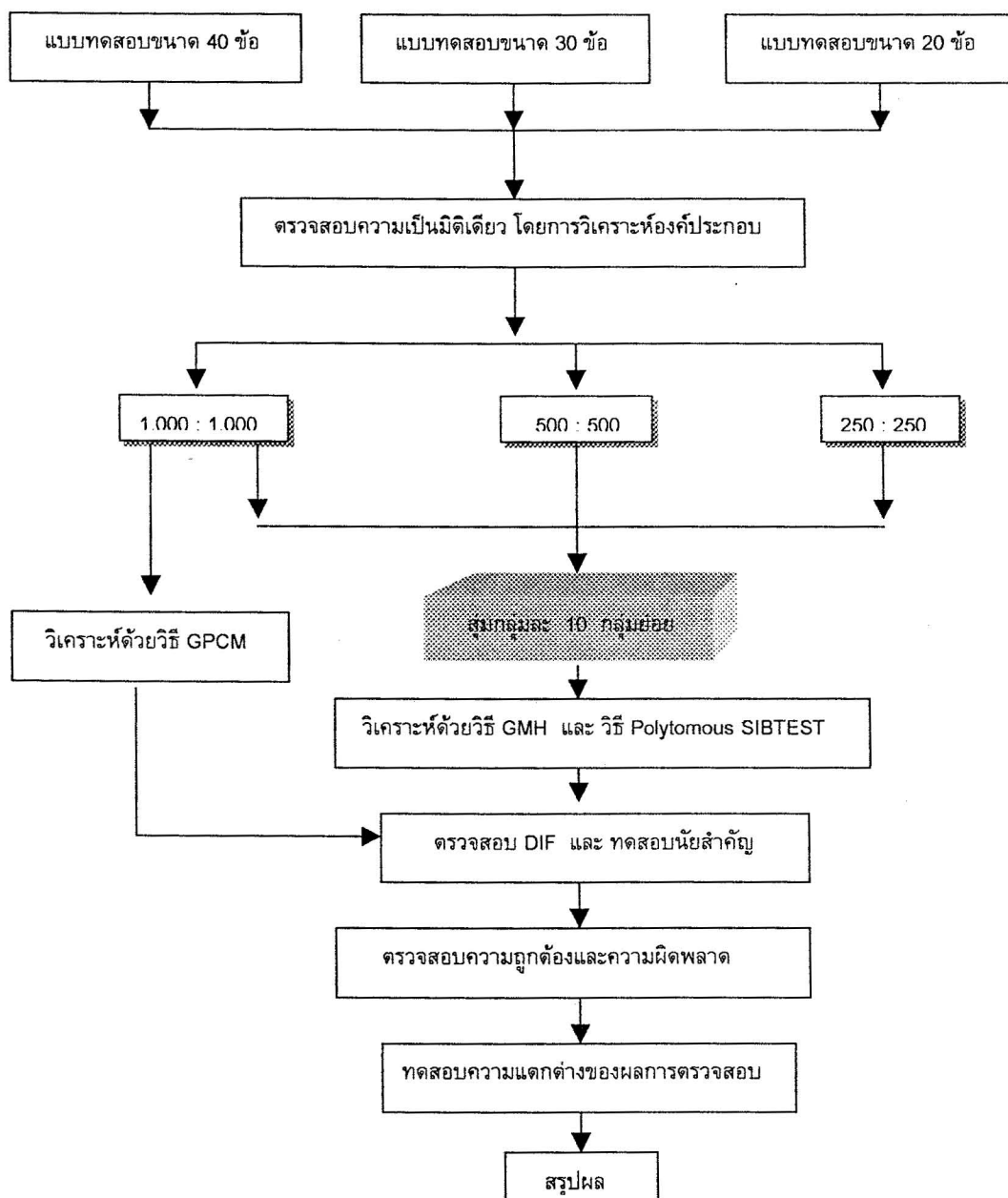
2.1 ข้อมูลของกลุ่มตัวอย่างที่ใช้ในการวิเคราะห์ด้วยวิธี GPCM

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GPCM จะใช้เป็นวิธีมาตรฐานในการระบุว่าข้อสอบข้อใดแสดงการทำหน้าที่ต่างกัน เพื่อใช้เป็นเกณฑ์ในการประเมินประสิทธิภาพการตรวจสอบของวิธีอื่น โดยวิธี GPCM จะใช้วิเคราะห์ในเงื่อนไขที่กลุ่มตัวอย่างมีขนาด 1,000 คน กับ ความยาวของแบบทดสอบ 3 ขนาด คือ 40 ข้อ, 30 ข้อ และ 20 ข้อ ดังนั้นการเตรียมข้อมูลผลการตอบเพื่อวิเคราะห์ด้วยวิธีนี้จึงทำโดยการสุ่มกลุ่มตัวอย่างจากประชากรเทียมของแต่ละเขตเป็นกลุ่มตัวอย่างขนาด 1,000 คน เท่านั้น

2.2 ข้อมูลของกลุ่มตัวอย่างที่ใช้ในการวิเคราะห์ด้วยวิธี GMH และวิธี Polytomous SIBTEST

การเตรียมข้อมูลของกลุ่มตัวอย่างเพื่อใช้ในการวิเคราะห์ด้วยวิธี GMH และวิธี Polytomous SIBTEST กรณีที่แบบทดสอบมีความยาว 40 ข้อ ทำการสุ่มกลุ่มตัวอย่างจากประชากรเทียมของแต่ละเขตเป็นกลุ่มตัวอย่างที่มีขนาด 1,000 คน, 500 คน และ 250 คน โดยวิธีสุ่มแบบ ใส่คืนขนาดละ 10 กลุ่มทำซ้ำ เพื่อให้เกิดความคงเส้นคงวาในการตรวจสอบ โดยใช้โปรแกรมสำเร็จรูป SPSS for Windows แล้วบันทึกเป็นแฟ้มข้อมูลผลการตอบของแบบทดสอบขนาด 40 ข้อ จำนวน $2 \times 3 \times 10 = 60$ แฟ้ม (เขต $\times$ ขนาดกลุ่มตัวอย่าง $\times$ กลุ่มทำซ้ำ) ส่วนกรณีที่แบบทดสอบที่มีความยาว 30 ข้อ และ 20 ข้อ ดำเนินการเช่นเดียวกัน ได้แฟ้มข้อมูลผลการตอบของแบบทดสอบขนาด 30 ข้อ และ 20 ข้อ ขนาดละจำนวน 60 แฟ้ม รวมทั้งหมด 180 แฟ้ม

3. การเตรียมข้อมูลเพื่อใช้ในการวิเคราะห์ สรุปได้ดังภาพประกอบ 3



ภาพประกอบ 3 กระบวนการเตรียมข้อมูลเพื่อใช้ในการวิเคราะห์

การวิเคราะห์ข้อมูล

หลังจากดำเนินการจัดเตรียมข้อมูลแล้ว ผู้วิจัยได้ดำเนินการวิเคราะห์ข้อมูลตามขั้นตอน ดังนี้

1. การตรวจสอบความเป็นมิติเดียว (Unidimensionality) ของแบบทดสอบ ใช้การวิเคราะห์องค์ประกอบ (Factor Analysis) ด้วยการวิเคราะห์องค์ประกอบหลัก (Principle Component Analysis) หมุนแกนด้วยวิธีแวนริแมกซ์ (Varimax) ตามข้อเสนอของ ลอร์ด เพื่อตรวจสอบความเป็นมิติเดียวของแบบทดสอบ ซึ่งเป็นข้อตกลงเบื้องต้นของการวิเคราะห์การกำหนดหน้าที่ต่างกันของข้อสอบตามวิธีการของทฤษฎีการตอบข้อสอบ (IRT) โดยใช้โปรแกรมสำเร็จรูป SPSS for Window โดยใช้ข้อมูลผลการตอบ

ข้อสอบของนักเรียนกลุ่มตัวอย่างที่กำหนดให้เป็นประชากรเทียม จำนวน 2,247 คน จำแนกเป็นกลุ่มเพศชายจำนวน 1,025 คนและเพศหญิงจำนวน 1,222 คน เมื่อแบบทดสอบมีขนาด 40 ข้อ, 30 ข้อ และ 20 ข้อ

2. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GPCM

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในการศึกษาครั้งนี้มี 3 วิธี คือ วิธี GMH, วิธี Polytomous SIBTEST และวิธี GPCM โดยจะใช้วิธี GMH และวิธี Polytomous SIBTEST วิเคราะห์ในทุกเงื่อนไขของขนาดกลุ่มตัวอย่างและความยาวของแบบทดสอบ เพื่อเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) แต่วิธี GPCM จะใช้เป็นวิธีมาตรฐานในการระบุว่าข้อสอบข้อใดแสดงการทำหน้าที่ต่างกัน เพื่อใช้เป็นเกณฑ์ในการประเมินประสิทธิภาพการตรวจสอบของวิธีอื่น โดยวิธี GPCM จะใช้วิเคราะห์ในเงื่อนไขที่กลุ่มตัวอย่างเพศชายขนาด 1,000 คน ต่อกลุ่มตัวอย่างเพศหญิงขนาด 1,000 คนกับความยาวของแบบทดสอบ 3 ขนาด คือ 40 ข้อ, 30 ข้อและ 20 ข้อ เนื่องจากวิธี GPCM ซึ่งเป็นวิธีตามทฤษฎีการตอบข้อสอบ (IRT) จะตรวจสอบได้อย่างมีประสิทธิภาพเมื่อใช้กลุ่มตัวอย่างขนาดใหญ่ ดังนั้น วิธีการนี้จึงวิเคราะห์เพียง 3 เงื่อนไขดังกล่าว มีขั้นตอนการวิเคราะห์ ดังนี้

2.1. การประมาณค่าพารามิเตอร์ข้อสอบของกลุ่มตัวอย่างเพศชายและเพศหญิงกลุ่มละ 1,000 คน ที่สุ่มมาจากประชากรเทียมของแต่ละกลุ่มเพศ เมื่อแบบทดสอบมีขนาด 40 ข้อ, 30 ข้อและ 20 ข้อ โดยใช้โปรแกรม PARSCALE Version 2.2 พัฒนาขึ้นโดย มูรากิ และบอค (Muraki and Bock, 1991) ซึ่งใช้วิเคราะห์ข้อสอบที่ตรวจให้คะแนนแบบหลายค่า เพื่อประมาณค่าพารามิเตอร์ของข้อสอบ ตามแนวทฤษฎีการตอบข้อสอบ (IRT) ผลการวิเคราะห์จะได้แฟ้มข้อมูลของค่าประมาณพารามิเตอร์ข้อสอบของผู้สอบกลุ่มเพศชายและเพศหญิง

2.2 ปรับเทียบสเกลพารามิเตอร์ข้อสอบของทั้งสองกลุ่ม ด้วยการแปลงค่าเชิงเส้น (linear transformation) โดยใช้สูตรในการปรับเทียบ (Tate, R. L. 1999)

$$a_{jf}^* = a_{jf} / A$$

$$b_{jf}^* = A b_{jf} + K$$

- เมื่อ
- a_{jf}^* แทน ค่าอำนาจจำแนกของข้อสอบข้อที่ j จากผู้สอบกลุ่มสนใจที่แปลงค่าแล้ว
 - b_{jf}^* แทน ค่าความยากของข้อสอบข้อที่ j จากผู้สอบกลุ่มสนใจที่แปลงค่าแล้ว
 - a_{jf} แทน ค่าอำนาจจำแนกของข้อสอบข้อที่ j จากผู้สอบกลุ่มสนใจที่ยังไม่ได้แปลงค่า
 - b_{jf} แทน ค่าความยากของข้อสอบข้อที่ j จากผู้สอบกลุ่มสนใจที่ยังไม่ได้แปลงค่า
 - A แทน ค่าความชัน (slope)
 - K แทน ค่าจุดตัด (Intercept)

เนื่องจากพารามิเตอร์ของข้อสอบดังกล่าวประมาณค่ามาจากการวิเคราะห์แยกกันระหว่างกลุ่มผู้สอบย่อย ดังนั้น ก่อนที่จะนำข้อมูลของทั้งสองกลุ่มมาตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจะต้องปรับเทียบสเกลพารามิเตอร์ของข้อสอบให้อยู่บนสเกลเมตริกซ์เดียวกัน

2.3 นำผลการปรับเทียบสเกลพารามิเตอร์ในข้อ 2.2 มาตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ตามวิธี GPCM เพื่อใช้เป็นวิธีมาตรฐานในการระบุว่าข้อสอบข้อใดทำหน้าที่ต่างกัน ซึ่งสามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ 3 วิธี คือ วิธีการทดสอบไค-สแควร์ของลอร์ด (Lord's (1980) χ^2), วิธีการวัดพื้นที่สัมบูรณ์ของราชู (Exact Area Measures) (Raju, 1990) และวิธีการวัดพื้นที่ช่วงปิด

ของคิมและโคเฮน (Closed-Interval Area Measures) (Kim & Cohen, 1991) ภายใต้โมเดลโลจิสติกแบบ 1, 2, 3 พารามิเตอร์ ในการศึกษาค้นคว้าวิจัยใช้วิธีการทดสอบไค-สแควร์ของลอร์ด (1980) ซึ่งใช้สถิติทดสอบไค-สแควร์ (χ^2) โดยใช้โปรแกรมสำเร็จรูป MATLAB ช่วยในการวิเคราะห์

สถิติทดสอบ ไค-สแควร์ (χ^2) คำนวณได้จากสูตร

$$\chi_i^2 = V_i' \Gamma_i^{-1} V_i$$

โดยที่
$$V_i' = [\hat{a}_{iF} - \hat{a}_{iR}, \hat{b}_{iF_1} - \hat{b}_{iR_1}, \dots, \hat{b}_{iF_m} - \hat{b}_{iR_m}]$$

$$\Gamma_i = \Gamma_{iF} + \Gamma_{iR}$$

เมื่อ $\hat{a}_{iF}$ แทน ค่าประมาณพารามิเตอร์อำนาจจำแนกของข้อสอบข้อที่ i จากผู้สอบกลุ่มสนใจ

$\hat{a}_{iR}$ แทน ค่าประมาณพารามิเตอร์อำนาจจำแนกของข้อสอบข้อที่ i จากผู้สอบกลุ่มอ้างอิง

$\hat{b}_{iF_m}$ แทน ค่าประมาณพารามิเตอร์ความยากของข้อสอบข้อที่ i จากผู้สอบกลุ่มสนใจ
ในลำดับขั้นที่ m

$\hat{b}_{iR_m}$ แทน ค่าประมาณพารามิเตอร์ความยากของข้อสอบข้อที่ i จากผู้สอบกลุ่มอ้างอิง
ในลำดับขั้นที่ m

V_i' แทน เวกเตอร์ความแตกต่างของค่าประมาณพารามิเตอร์ของข้อสอบข้อที่ i
จากผู้สอบกลุ่มสนใจและกลุ่มอ้างอิง

Γ_i แทน เมตริกซ์ความแปรปรวน-ความแปรปรวนร่วมของ $(\hat{a}_{iF} - \hat{a}_{iR})$ และ

$(\hat{b}_{iF_m} - \hat{b}_{iR_m})$ ที่มีขนาด $m \times m$

Γ_i^{-1} แทน อินเวอร์สของเมตริกซ์ Γ_i

3. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GMH

นำข้อมูลผลการตอบกลุ่มเพศชายและเพศหญิง ในเงื่อนไขความยาวของแบบทดสอบ และ กลุ่มตัวอย่างขนาดเดียวกันแต่ละกลุ่มมาดำเนินการวิเคราะห์ เพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GMH โดยจับคู่นักเรียนเพศชายและเพศหญิงที่มีความสามารถเท่ากัน ซึ่งพิจารณาจากคะแนนรวมของผลการตอบข้อสอบวิชาคณิตศาสตร์ที่เท่ากัน แล้วนำข้อมูลผลการตอบของนักเรียนที่จับคู่แล้วทั้ง 2 กลุ่มมาสร้างตารางการถัว 2xTxK เมื่อ T เป็นจำนวนลำดับขั้นการตอบ และ k เป็นจำนวนของระดับของการแบ่งชั้นตัวแปรการจับคู่ ในแต่ละระดับทั้งหมด k ระดับ เพื่อหาอัตราส่วนแต่มีต่อรวม α และใช้สถิติ GMH_{χ^2} ทดสอบนัยสำคัญ โดยใช้โปรแกรมสำเร็จรูป EXCEL และ MATLAB ช่วยในการวิเคราะห์

การทดสอบนัยสำคัญของสมมติฐาน ใช้สถิติ GMH_{χ^2} คำนวณได้จากสูตร

$$GMH_{\chi^2}^2 = [\sum A_k - \sum E(A_k)]' [\sum V(A_k)]^{-1} [\sum A_k - \sum E(A_k)] , \quad df = T-1$$

$$\begin{aligned}
\text{โดยที่ } A'_k &= (n_{R1k}, n_{R2k}, \dots, n_{R(T-1)k}) \\
E(A'_k) &= n_{R+k} n'_k / n_{++k} \\
n'_k &= (n_{+1k}, n_{+2k}, \dots, n_{+(T-1)k}) \\
V(A_k) &= n_{R+k} n_{F+k} \left(\frac{n_{++k} \text{diag}(n_k) - n_k n'_k}{n_{++k}^2 (n_{++k} - 1)} \right)
\end{aligned}$$

และ $\text{diag}(n_k)$ เป็นเมตริกซ์แนวทแยง (diagonal matrix) มิติ $(T-1) \times (T-1)$ ที่มีสมาชิกเป็น n_k
เมื่อ A_k และ $E(A_k)$ เป็นเวกเตอร์ที่ยาว $T-1$ ที่สมนัยกับ $T-1$ ไต ๆ ของลำดับชั้นการตอบ T
และ $V(A_k)$ เป็นเมตริกซ์ความแปรปรวน-ความแปรปรวนร่วม (Variance-covariance matrix) มิติ $(T-1) \times (T-1)$

4. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วย วิธี Polytomous SIBTEST

นำข้อมูลของกลุ่มเพศชายและเพศหญิงในเงื่อนไขความยาวของแบบทดสอบและกลุ่มตัวอย่างขนาดเดียวกัน แต่ละคู่มาดำเนินการวิเคราะห์ เพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี Polytomous SIBTEST ด้วยโปรแกรมสำเร็จรูป DIFPACK โปรแกรมย่อย Poly-SIBTEST

5. คำนวณประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

หลังจากวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GMH และวิธี Polytomous SIBTEST แล้วนำผลการวิเคราะห์ที่ได้มาหาความถูกต้องของการตรวจสอบและความผิดพลาด ด้วยการเปรียบเทียบกับผลที่ได้จากการตรวจสอบด้วยวิธี GPCM โดยจะใช้เกณฑ์ว่า ข้อสอบที่ถูกตรวจสอบพบว่าทำหน้าที่ต่างกันด้วยวิธี GPCM ถือว่าเป็นข้อที่ทำหน้าที่ต่างกัน ถ้าวิธี GMH และวิธี Polytomous SIBTEST ตรวจสอบพบข้อสอบที่ทำหน้าที่ต่างกันตรงกับข้อสอบที่ถูกตรวจสอบพบด้วยวิธี GPCM ถือว่าวิธีการนั้นมีความถูกต้องในการตรวจสอบ แต่ถ้าการตรวจสอบพบว่าข้อสอบทำหน้าที่ต่างกันไม่ตรงกับข้อสอบที่ตรวจพบด้วยวิธี GPCM ถือว่า มีความผิดพลาดในการตรวจสอบ ซึ่งการคำนวณความถูกต้องในการตรวจสอบและความผิดพลาดในการตรวจสอบ เป็นดังนี้

ความถูกต้องในการตรวจสอบ (Power of Test)

$$= \frac{\text{จำนวนข้อสอบที่ระบุว่าทำหน้าที่ต่างกันได้ถูกต้อง}}{\text{จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งหมดในแบบทดสอบ}}$$

ความผิดพลาดในการตรวจสอบ (False Positive : FP)

$$= \frac{\text{จำนวนข้อสอบที่ระบุผิดพลาดว่าทำหน้าที่ต่างกัน}}{\text{จำนวนข้อสอบที่ทำหน้าที่ไม่ต่างกันทั้งหมดในแบบทดสอบ}}$$

6. การทดสอบความแตกต่างของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีการตรวจสอบที่มีผลต่อความถูกต้องและความผิดพลาดในการตรวจสอบ โดยใช้เทคนิคการวิเคราะห์ความแปรปรวนแบบหลายตัวแปร (Multivariate analysis of variance : MANOVA) โดยใช้โปรแกรมสำเร็จรูป SPSS for Windows

7. เกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี GMH และวิธี Polytomous SIBTEST จากการที่ผู้วิจัยใช้ผลจากวิธีการตรวจสอบด้วยวิธี GPCM เป็นเกณฑ์ในการเปรียบเทียบ เพื่อคำนวณหาความถูกต้องในการตรวจสอบ และ ความผิดพลาดในการตรวจสอบของแต่ละวิธีนั้น ผู้วิจัยกำหนดเกณฑ์ที่ใช้เปรียบเทียบประสิทธิภาพในการทำหน้าที่ต่างกันของข้อสอบของวิธีการตรวจสอบทั้ง 2 วิธี คือ วิธีการใดมีความถูกต้องในการตรวจสอบ สูง และมีความผิดพลาดประเภทที่ 1 ต่ำ แสดงว่า วิธีการนั้นมีประสิทธิภาพสูงสุดในการทำหน้าที่ต่างกันของข้อสอบ

บทที่ 4

ผลการวิเคราะห์ข้อมูล

การวิจัยครั้งนี้ มีจุดมุ่งหมายเพื่อเปรียบเทียบประสิทธิภาพของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในแบบทดสอบที่มีการให้คะแนนแบบหลายค่าระหว่างวิธี GMH และ วิธี Polytomous SIBTEST เมื่อใช้ขนาดกลุ่มตัวอย่างและความยาวของแบบทดสอบต่างกัน ซึ่งผลการวิเคราะห์ข้อมูลครั้งนี้ผู้วิจัยได้เสนอเรียงตามลำดับ ดังนี้

ตอนที่ 1 การวิเคราะห์ข้อมูลเบื้องต้น

ตอนที่ 2 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GPCM โดยเสนอค่าพารามิเตอร์ของข้อสอบ และผลของการตรวจสอบ

ตอนที่ 3 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 2 วิธี ได้แก่ วิธี GMH และวิธี Polytomous SIBTEST โดยเรียงลำดับดังนี้

3.1 เปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี GMH และวิธี Polytomous SIBTEST

3.2 เปรียบเทียบความถูกต้องของการตรวจสอบระหว่างวิธี GMH และวิธี Polytomous SIBTEST

3.3 เปรียบเทียบความผิดพลาดของการตรวจสอบระหว่างวิธี GMH และวิธี Polytomous SIBTEST

เพื่อให้การนำเสนอผลการวิเคราะห์ข้อมูลมีความเข้าใจที่ตรงกันและง่ายขึ้น ผู้วิจัยจึงกำหนดสัญลักษณ์แทน ดังนี้

N	แทน จำนวนคน
I	แทน จำนวนข้อสอบ
$\bar{X}$	แทน ค่าเฉลี่ยเลขคณิต
SD	แทน ค่าส่วนเบี่ยงเบนมาตรฐาน
Max	แทน ค่าคะแนนสูงสุด
Min	แทน ค่าคะแนนต่ำสุด
CV	แทน สัมประสิทธิ์ความแปรผัน (Coefficient of Variation)
α	แทน ค่าความเชื่อมั่นของแบบทดสอบ
a_i	แทน ค่าพารามิเตอร์ความชัน (Slope parameter) หรือค่าอำนาจจำแนกของข้อสอบ รวม
b_i	แทน ค่าพารามิเตอร์ลำดับชั้นข้อสอบ (Item-category parameter) หรือค่าความยากของข้อสอบ รวม
a_R	แทน ค่าพารามิเตอร์ความชัน (Slope parameter) หรือค่าอำนาจจำแนกของข้อสอบ ของกลุ่มเพศชาย

b_{R1}	แทน ค่าพารามิเตอร์ลำดับชั้นข้อสอบ (Item-category parameter) หรือค่าความยากของข้อสอบ ของกลุ่มเพศชาย ลำดับชั้นที่ 1
b_{R2}	แทน ค่าพารามิเตอร์ลำดับชั้นข้อสอบ (Item-category parameter) หรือค่าความยากของข้อสอบ ของกลุ่มเพศชาย ลำดับชั้นที่ 2
b_{R3}	แทน ค่าพารามิเตอร์ลำดับชั้นข้อสอบ (Item-category parameter) หรือค่าความยากของข้อสอบ ของกลุ่มเพศชาย ลำดับชั้นที่ 3
b_{R4}	แทน ค่าพารามิเตอร์ลำดับชั้นข้อสอบ (Item-category parameter) หรือค่าความยากของข้อสอบ ของกลุ่มเพศชาย ลำดับชั้นที่ 4
b_{R5}	แทน ค่าพารามิเตอร์ลำดับชั้นข้อสอบ (Item-category parameter) หรือค่าความยากของข้อสอบ ของกลุ่มเพศชาย ลำดับชั้นที่ 5
a_F	แทน ค่าพารามิเตอร์ความชัน (Slope parameter) หรือค่าอำนาจจำแนกของข้อสอบ ของกลุ่มเพศหญิง
b_{F1}	แทน ค่าพารามิเตอร์ลำดับชั้นข้อสอบ (Item-category parameter) หรือค่าความยากของข้อสอบ ของกลุ่มเพศหญิง ลำดับชั้นที่ 1
b_{F2}	แทน ค่าพารามิเตอร์ลำดับชั้นข้อสอบ (Item-category parameter) หรือค่าความยากของข้อสอบ ของกลุ่มเพศหญิง ลำดับชั้นที่ 2
b_{F3}	แทน ค่าพารามิเตอร์ลำดับชั้นข้อสอบ (Item-category parameter) หรือค่าความยากของข้อสอบ ของกลุ่มเพศหญิง ลำดับชั้นที่ 3
b_{F4}	แทน ค่าพารามิเตอร์ลำดับชั้นข้อสอบ (Item-category parameter) หรือค่าความยากของข้อสอบ ของกลุ่มเพศหญิง ลำดับชั้นที่ 4
b_{F5}	แทน ค่าพารามิเตอร์ลำดับชั้นข้อสอบ (Item-category parameter) หรือค่าความยากของข้อสอบ ของกลุ่มเพศหญิง ลำดับชั้นที่ 5
DIF	แทน การทำหน้าที่ต่างกันของข้อสอบ
GPCM	แทน วิธีการ Generalized Partial Credit Model ของทฤษฎีการตอบข้อสอบ (IRT)
GMH	แทน วิธีการ Generalized Mantel-Haenszel
Poly-SIB	แทน วิธีการ Polytomous SIBTEST
P	แทน ความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ
E	แทน ความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ
A	แทน ค่าความชัน (Slope) ของเส้นถดถอย
K	แทน ค่าจุดตัด (Intercept) บนแกนตั้ง
*	แทน $p < .05$
**	แทน $p < .01$

ตอนที่ 1 การวิเคราะห์ข้อมูลเบื้องต้น

1.1 ค่าสถิติพื้นฐานของข้อมูลผลการสอบและคุณภาพของแบบทดสอบ

การวิเคราะห์ข้อมูลส่วนนี้ผู้วิจัยได้นำคะแนนที่ได้จากการตอบข้อสอบของแบบทดสอบวิชาคณิตศาสตร์ เรื่อง วงกลม จำนวน 40 ข้อ, 30 ข้อ และ 20 ข้อ คำนวณค่าสถิติพื้นฐาน โดยแบบทดสอบฉบับ 30 ข้อ และ 20 ข้อ ได้มาโดยการคัดเลือกตามสัดส่วนในตารางกำหนดข้อสอบ (ตาราง 7) ปรากฏผลดังตาราง 8 และคำนวณหาคุณภาพของแบบทดสอบ ปรากฏผลดังแสดงในภาคผนวก จ

ตาราง 8 ค่าสถิติพื้นฐานของแบบทดสอบวิชาคณิตศาสตร์ ขนาด 40 ข้อ, 30 ข้อ และ 20 ข้อ

ค่าสถิติ	ความยาวของแบบทดสอบ		
	40 ข้อ	30 ข้อ	20 ข้อ
คะแนนเต็ม	200	150	100
ค่าคะแนนสูงสุด (Max)	193	147	100
ค่าคะแนนต่ำสุด (Min)	82	56	36
ค่าคะแนนเฉลี่ย ($\bar{X}$)	116.71	87.62	58.67
ค่าส่วนเบี่ยงเบนมาตรฐาน (SD)	17.23	13.99	10.81
สัมประสิทธิ์ความแปรผัน (CV)	14.76%	15.97%	18.43%
ความเบ้ (Skewness)	1.60	1.49	1.32
ความโด่ง (Kurtosis)	3.44	3.04	2.08
ค่าความเชื่อมั่น	.8462	.8256	.8045
จำนวนผู้สอบ(N)	2247	2247	2247

ผลการวิเคราะห์ตามตาราง 8 พบว่า แบบทดสอบขนาด 40 ข้อ มีคะแนนเต็ม 200 คะแนน ค่าคะแนนสูงสุด 193 คะแนน ค่าคะแนนต่ำสุด 82 คะแนน ค่าเฉลี่ยของคะแนนเท่ากับ 116.71 ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 17.23 ค่าสัมประสิทธิ์ความแปรผันเท่ากับ 14.76% ค่าความเบ้เท่ากับ 1.60 ค่าความโด่งเท่ากับ 3.44 และค่าความเชื่อมั่นของแบบทดสอบ เท่ากับ .8462

แบบทดสอบขนาด 30 ข้อ มีคะแนนเต็ม 150 คะแนน ค่าคะแนนสูงสุด 147 คะแนน ค่าคะแนนต่ำสุด 56 คะแนน ค่าเฉลี่ยของคะแนนเท่ากับ 87.62 ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 13.99 ค่าสัมประสิทธิ์ความแปรผันเท่ากับ 15.79% ค่าความเบ้เท่ากับ 1.49 ค่าความโด่งเท่ากับ 3.04 และค่าความเชื่อมั่นของแบบทดสอบ เท่ากับ .8256

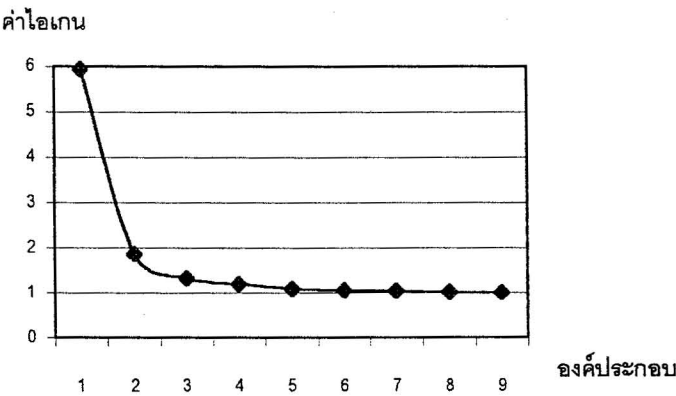
แบบทดสอบขนาด 20 ข้อ มีคะแนนเต็ม 100 คะแนน ค่าคะแนนสูงสุด 100 คะแนน ค่าคะแนนต่ำสุด 36 คะแนน ค่าเฉลี่ยของคะแนนเท่ากับ 58.67 ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10.81 ค่าสัมประสิทธิ์ความแปรผันเท่ากับ 18.43% ค่าความเบ้เท่ากับ 1.32 ค่าความโด่งเท่ากับ 2.08 และค่าความเชื่อมั่นของแบบทดสอบ เท่ากับ .8045

1.2 การตรวจสอบความเป็นมิติเดียวของแบบทดสอบ

การตรวจสอบความเป็นมิติเดียวของแบบทดสอบ เป็นข้อตกลงพื้นฐานที่สำคัญของทฤษฎีการตอบข้อสอบ IRT เป็นการตรวจสอบว่าแบบทดสอบวัดคุณลักษณะแฝงเพียงลักษณะเดียว (single latent trait) ผู้วิจัยได้นำคะแนนที่ได้จากการตอบข้อสอบวิชาคณิตศาสตร์ เรื่องวงกลม จำนวน 40 ข้อ, 30 ข้อและ 20 ข้อ มาตรวจสอบความเป็นมิติเดียว โดยใช้ การวิเคราะห์องค์ประกอบ (Factor Analysis) ด้วยการวิเคราะห์องค์ประกอบหลัก (Principle Component Analysis) หมุนแกนด้วยวิธีแวนิแมกซ์ (Varimax) หาค่าไอเกน (Eigen Value) และร้อยละของความแปรปรวน ปรากฏผลดังตาราง 9, 10 และ 11 และภาพประกอบ 4, 5 และ 6

ตาราง 9 ค่าไอเกน และร้อยละของความแปรปรวนขององค์ประกอบ ของแบบทดสอบขนาด 40 ข้อ

องค์ประกอบ	ค่าไอเกน	ร้อยละของความแปรปรวน
1	5.943	14.856
2	1.867	4.669
3	1.336	3.340
4	1.199	2.998
5	1.094	2.734
6	1.060	2.649
7	1.046	2.615
8	1.026	2.564
9	1.004	2.509



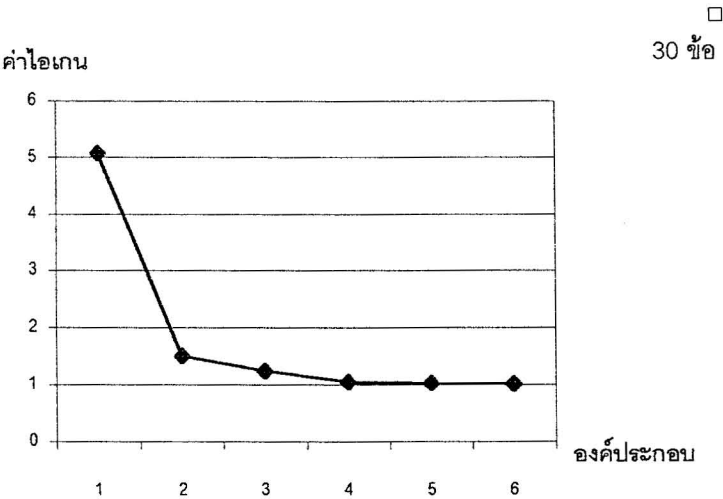
ภาพประกอบ 4 ผลการตรวจสอบความเป็นมิติเดียวของแบบทดสอบขนาด 40 ข้อ

ผลการวิเคราะห์ตามตาราง 9 พบว่า แบบทดสอบขนาด 40 ข้อ ค่าไอเกนขององค์ประกอบที่ 1 มีค่าสูงสุด(5.943) ซึ่งมีค่าสูงกว่าค่าไอเกนขององค์ประกอบที่ 2(1.867) ประมาณ 3.183 เท่า องค์ประกอบที่ 1 มีค่าไอเกนแตกต่างจากค่าไอเกนขององค์ประกอบอื่นอย่างเด่นชัด ส่วนค่าไอเกนขององค์ประกอบอื่น ๆ มี

ค่าใกล้เคียงกัน และองค์ประกอบที่ 1 มีค่าร้อยละของความแปรปรวนค่อนข้างสูง (14.856) ซึ่งใกล้เคียงกับเกณฑ์ของเรคแคส (Reckase. อ้างถึงใน Raju. 1993 : 304) ที่เสนอว่า ค่าร้อยละของความแปรปรวนควรมีค่าไม่น้อยกว่าร้อยละ 20 จึงจะถือว่าแบบทดสอบมีความเป็นมิติตีเดียว เมื่อพิจารณาค่าไอเกนและค่าร้อยละของความแปรปรวนขององค์ประกอบที่ 1 และ ภาพประกอบ 4 ประกอบกันแล้ว ในทางปฏิบัติถือได้ว่าแบบทดสอบฉบับนี้มีความเป็นมิติตีเดียว ดังนั้นจึงสามารถนำข้อมูลไปวิเคราะห์ตามทฤษฎีการตอบข้อสอบได้

ตาราง 10 ค่าไอเกน และร้อยละของความแปรปรวนขององค์ประกอบ ของแบบทดสอบขนาด 30 ข้อ

องค์ประกอบ	ค่าไอเกน	ร้อยละของความแปรปรวน
1	5.082	16.939
2	1.502	5.006
3	1.236	4.119
4	1.049	3.497
5	1.028	3.427
6	1.002	3.339

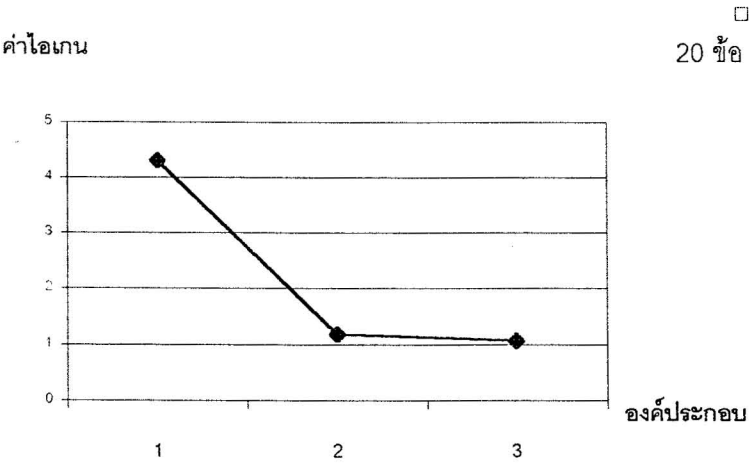


ภาพประกอบ 5 ผลการตรวจสอบความเป็นมิติตีเดียวของแบบทดสอบขนาด 30 ข้อ

ผลการวิเคราะห์ตามตาราง 10 พบว่า แบบทดสอบขนาด 30 ข้อ ค่าไอเกนขององค์ประกอบที่ 1 มีค่าสูงสุด(5.082) ซึ่งมีค่าสูงกว่าค่าไอเกนขององค์ประกอบที่ 2 (1.502) ประมาณ 3.383 เท่า องค์ประกอบที่ 1 มีค่าไอเกนแตกต่างจากค่าไอเกนขององค์ประกอบอื่นอย่างเด่นชัด ส่วนค่าไอเกนขององค์ประกอบอื่น ๆ มีค่าใกล้เคียงกัน และค่าร้อยละของความแปรปรวนขององค์ประกอบที่ 1 มีค่าค่อนข้างสูง (16.939) เมื่อพิจารณาค่าไอเกนและค่าร้อยละของความแปรปรวนขององค์ประกอบที่ 1 และภาพประกอบ 5 ประกอบกันแล้ว ในทางปฏิบัติถือได้ว่าแบบทดสอบฉบับนี้มีความเป็นมิติตีเดียว ดังนั้น จึงสามารถนำข้อมูลไปวิเคราะห์ตามทฤษฎีการตอบข้อสอบได้

ตาราง 11 ค่าไอเกน และร้อยละของความแปรปรวนขององค์ประกอบ ของแบบทดสอบขนาด 20 ข้อ

องค์ประกอบ	ค่าไอเกน	ร้อยละของความแปรปรวน
1	4.305	21.523
2	1.179	5.894
3	1.068	5.341



ภาพประกอบ 6 ผลการตรวจสอบความเป็นมิติเดียวของแบบทดสอบขนาด 20 ข้อ

ผลการวิเคราะห์ตามตาราง 11 พบว่า แบบทดสอบขนาด 20 ข้อ ค่าไอเกนขององค์ประกอบที่ 1 มีค่าสูงสุด (4.305) ซึ่งมีค่าสูงกว่าค่าไอเกนขององค์ประกอบที่ 2 (1.179) ประมาณ 3.651 เท่า องค์ประกอบที่ 1 มีค่าไอเกนแตกต่างจากค่าไอเกนขององค์ประกอบอื่นอย่างเด่นชัด ส่วนค่าไอเกนขององค์ประกอบอื่นๆ มีค่าใกล้เคียงกัน และองค์ประกอบที่ 1 มีค่าร้อยละของความแปรปรวนค่อนข้างสูง (21.523) เมื่อพิจารณา ค่าไอเกนและค่าร้อยละของความแปรปรวนขององค์ประกอบที่ 1 และ ภาพประกอบ 6 ประกอบกันแล้ว ในทางปฏิบัติถือว่าแบบทดสอบฉบับนี้มีความเป็นมิติเดียว ดังนั้น จึงสามารถนำข้อมูลไปวิเคราะห์ตามทฤษฎีการตอบข้อสอบได้

ตอนที่ 2 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GPCM

หลังจากตรวจสอบความเป็นมิติเดียวของข้อมูล ซึ่งเป็นข้อตกลงเบื้องต้นในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบตามแนวคิดของทฤษฎีการตอบข้อสอบ (IRT) แล้ว พบว่า แบบทดสอบขนาด 40 ข้อ, 30 ข้อ และ 20 ข้อ มีความเป็นมิติเดียว จึงตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GPCM โดยการประมาณค่าพารามิเตอร์ของข้อสอบของกลุ่มเพศชายและกลุ่มเพศหญิง ขนาดกลุ่มละ 1,000 คน เมื่อความยาวของแบบทดสอบมีขนาด 40 ข้อ, 30 ข้อ และ 20 ข้อ โดยใช้โปรแกรม PARSCALE version 2.2 วิเคราะห์ค่าพารามิเตอร์ความชัน(Slope parameter) หรือค่าอำนาจจำแนกของข้อสอบ(a) และค่าพารามิเตอร์ ลำดับชั้นของข้อสอบ (Item-category parameter) หรือ ค่าความยากลำดับชั้นของข้อสอบ (b) ปรากฏผลดัง ตาราง 12 - 14

ตาราง 12 ค่าพารามิเตอร์ของข้อสอบ แยกตามกลุ่มเพศชาย – หญิง เมื่อกลุ่มตัวอย่าง 1,000 คนและ
ความยาวของแบบทดสอบ 40 ข้อ

ข้อที่	ค่าพารามิเตอร์ของข้อสอบ											
	a _R	a _F	b _{R1}	b _{R2}	b _{R3}	b _{R4}	b _{R5}	b _{F1}	b _{F2}	b _{F3}	b _{F4}	b _{F5}
1	.314	.290	-3.56	-1.51	0.09	1.58	3.17	-3.39	-1.59	0.02	1.37	2.64
2	.235	.272	-4.16	-2.11	-0.50	0.98	2.58	-3.77	-1.98	-0.36	0.98	2.26
3	.301	.287	-3.71	-1.66	-0.06	1.42	3.02	-3.38	-1.59	0.23	1.37	2.65
4	.296	.388	-3.97	-1.92	-0.32	1.17	2.76	-3.96	-2.16	-0.55	0.79	2.07
5	.266	.301	-5.88	-3.83	-2.23	-0.74	0.85	-5.65	-3.85	-2.24	-0.89	0.39
6	.291	.321	-3.58	-1.52	0.08	1.56	3.16	-3.61	-1.82	-0.20	1.14	2.42
7	.246	.310	-3.63	-1.57	0.03	1.51	3.11	-3.36	-1.57	0.05	1.39	2.67
8	.292	.322	-3.24	-1.19	0.41	1.90	3.49	-3.08	-1.29	0.33	1.67	2.95
9	.239	.323	-4.23	-2.17	-0.57	0.91	2.51	-4.30	-2.51	-0.90	0.45	1.73
10	.271	.234	-1.15	-2.10	-0.50	0.99	2.58	-4.34	-2.54	-0.93	0.41	1.69
11	.287	.358	-4.00	-1.95	-0.35	1.13	2.73	-3.67	-1.88	-0.26	1.08	2.36
12	.273	.312	-4.71	-2.66	-1.06	0.42	2.02	-4.24	-2.45	-0.84	0.51	1.79
13	.211	.273	-4.51	-2.46	-0.86	0.63	2.22	-4.24	-2.44	-0.83	0.52	1.80
14	.278	.340	-4.77	-2.72	-1.11	0.37	1.96	-4.62	-2.83	-1.21	0.13	1.41
15	.316	.387	-4.58	-2.53	-0.92	0.56	2.15	-4.23	-2.44	-0.83	0.52	1.80
16	.313	.305	-3.72	-1.66	-0.06	1.42	3.02	-3.69	-1.90	-0.29	1.06	2.34
17	.357	.341	-3.98	-1.92	-0.32	1.16	2.76	-3.47	-1.68	-0.06	1.28	2.56
18	.321	.401	-3.74	-1.69	-0.09	1.40	2.99	-3.55	-1.75	-0.14	1.21	2.49
19	.424	.463	-4.65	-2.60	-1.00	0.48	2.08	-4.63	-2.83	-1.22	0.12	1.40
20	.252	.270	-5.73	-3.68	-2.08	-0.59	1.00	-5.65	-3.86	-2.24	-0.90	0.38
21	.298	.321	-3.98	-1.93	-0.33	1.15	2.75	-3.91	-2.11	-0.50	0.84	2.12
22	.326	.311	-3.55	-1.50	0.10	1.58	3.18	-3.24	-1.44	0.17	1.51	2.79
23	.410	.391	-4.53	-2.48	-0.88	0.60	2.20	-4.65	-2.85	-1.24	0.11	1.38
24	.278	.344	-4.73	-2.68	-1.08	0.40	2.00	-4.70	-2.90	-1.29	0.05	1.33
25	.253	.271	-3.95	-1.90	-0.30	1.19	2.78	-3.60	-1.80	-0.19	1.16	2.44
26	.308	.303	-3.85	-1.79	-0.19	1.29	2.89	-3.64	-1.84	-0.23	1.12	2.40
27	.275	.262	-3.64	-1.58	0.02	1.50	2.10	-3.65	-1.86	-0.24	1.10	2.38
28	.277	.324	-4.26	-2.20	-0.60	0.88	2.48	-4.10	-2.31	-0.69	0.65	1.93
29	.309	.349	-3.57	-1.52	0.08	1.56	3.16	-3.32	-1.53	0.09	1.43	2.71
30	.293	.368	-4.46	-2.41	-0.81	0.68	2.27	-4.66	-2.86	-1.25	0.09	1.37
31	.209	.250	-5.29	-3.24	-1.64	-0.16	1.44	-4.66	-2.86	-1.25	0.10	1.37
32	.317	.358	-4.49	-2.44	-0.84	0.64	2.24	-4.03	-2.24	-0.63	0.72	2.00
33	.260	.379	-5.04	-2.98	-1.38	0.10	1.70	-4.30	-2.50	-0.89	0.46	1.74
34	.320	.307	-4.61	-2.56	-0.96	0.53	2.12	-4.52	-2.73	-1.11	0.23	1.51
35	.351	.507	-4.52	-2.46	-0.86	0.62	2.22	-4.36	-2.57	-0.95	0.39	1.67

ตาราง 12 (ต่อ)

ข้อที่	ค่าพารามิเตอร์ของข้อสอบ											
	a _R	a _F	b _{R1}	b _{R2}	b _{R3}	b _{R4}	b _{R5}	b _{F1}	b _{F2}	b _{F3}	b _{F4}	b _{F5}
36	.286	.300	-5.71	-3.66	-2.06	-0.57	1.02	-5.76	-3.96	-2.35	-1.00	0.28
37	.236	.339	-5.76	-3.71	-2.11	-0.63	0.97	-5.81	-4.02	-2.40	-1.06	0.22
38	.299	.302	-3.98	-1.93	-0.33	1.16	2.75	-3.92	-2.12	-0.51	0.84	2.12
39	.270	.368	-4.01	-1.96	-0.36	1.13	2.72	-3.83	-2.03	-0.42	0.93	2.20
40	.338	.334	-3.80	-1.75	-0.15	1.33	2.93	-3.74	-1.94	-0.33	1.01	2.29
ค่าเฉลี่ย	0.29	0.33	-4.31	-2.25	-0.65	0.83	2.43	-4.13	-2.34	-0.72	0.62	1.90

ผลการวิเคราะห์ตามตาราง 12 พบว่า ค่าพารามิเตอร์ของข้อสอบของกลุ่มตัวอย่างเพศชายและเพศหญิง เมื่อกลุ่มตัวอย่างขนาด 1,000 คน และความยาวของแบบทดสอบขนาด 40 ข้อ กลุ่มตัวอย่างเพศชาย มีค่าอำนาจจำแนกเฉลี่ยเท่ากับ 0.29 และค่าความยากเฉลี่ยในแต่ละลำดับขั้นเท่ากับ -4.31 -2.25 -0.65 0.83 และ 2.43 ส่วนในกลุ่มตัวอย่างเพศหญิง มีค่าอำนาจจำแนกเฉลี่ยเท่ากับ 0.33 และค่าความยากเฉลี่ยในแต่ละลำดับขั้นเท่ากับ -4.13 -2.34 -0.72 0.62 และ 1.90

ตาราง 13 ค่าพารามิเตอร์ของข้อสอบ แยกตามกลุ่มเพศชาย – หญิง เมื่อกลุ่มตัวอย่าง 1000 คนและความยาวของแบบทดสอบ 30 ข้อ

ข้อที่	ค่าพารามิเตอร์ของข้อสอบ											
	a _R	a _F	b _{R1}	b _{R2}	b _{R3}	b _{R4}	b _{R5}	b _{F1}	b _{F2}	b _{F3}	b _{F4}	b _{F5}
1	.301	.309	-3.71	-1.95	-0.31	1.10	2.53	-3.49	-1.83	-0.37	0.96	2.04
2	.296	.366	-4.39	-2.63	-0.99	0.42	1.85	-3.98	-2.32	-0.86	0.47	1.55
3	.335	.336	-3.44	-1.68	-0.04	1.37	2.80	-3.40	-1.74	-0.28	1.05	2.13
4	.332	.403	-3.46	-1.70	-0.06	1.35	2.78	-3.31	-1.66	-0.19	1.14	2.22
5	.250	.311	-5.40	-3.64	-2.00	-0.59	0.84	-5.20	-3.54	-2.08	-0.75	0.33
6	.312	.356	-3.69	-1.93	-0.29	1.12	2.55	-3.60	-1.95	-0.48	0.85	1.92
7	.329	.366	-3.27	-1.52	0.13	1.53	2.96	-2.94	-1.29	0.18	1.51	2.58
8	.446	.429	-4.22	-2.46	-0.82	0.59	2.02	-4.25	-2.59	-1.12	0.21	1.28
9	.292	.302	-4.41	-2.65	-1.01	0.40	1.83	-4.45	-2.80	-1.33	-0.00	1.07
10	.321	.313	-3.56	-1.80	-0.16	1.25	2.68	-3.39	-1.73	-0.26	1.07	2.14
11	.287	.252	-3.36	-1.60	0.04	1.45	2.88	-3.39	-1.73	-0.27	1.07	2.14
12	.296	.369	-3.95	-2.19	-0.55	0.86	2.29	-3.68	-2.02	-0.56	0.77	1.85
13	.333	.374	-3.30	-1.54	0.10	1.51	2.94	-3.04	-1.38	0.08	1.41	2.49
14	.310	.419	-4.15	-2.19	-0.75	0.66	2.09	-4.28	-2.62	-1.16	0.18	1.25
15	.219	.290	-4.94	-3.19	-1.54	-0.14	1.30	-4.37	-2.72	-1.25	0.08	1.15
16	.310	.347	-4.19	-2.43	-0.79	0.62	2.05	-3.69	-2.03	-0.57	0.76	1.84
17	.273	.405	-4.70	-2.94	-1.30	0.11	1.54	-4.04	-2.38	-0.92	0.42	1.49
18	.349	.547	-4.21	-2.45	-0.81	0.60	2.03	-4.07	-2.41	-0.95	0.38	1.46

ตาราง 13 (ต่อ)

ข้อที่	ค่าพารามิเตอร์ของข้อสอบ											
	a _R	a _F	b _{R1}	b _{R2}	b _{R3}	b _{R4}	b _{R5}	b _{F1}	b _{F2}	b _{F3}	b _{F4}	b _{F5}
19	.315	.294	-5.32	-3.56	-1.92	-0.51	0.92	-5.45	-3.79	-2.33	-1.00	0.08
20	.248	.315	-5.39	-3.63	-1.99	-0.58	0.85	-5.56	-3.90	-2.44	-1.11	-0.03
21	.309	.372	-3.69	-1.93	-0.29	1.12	2.55	-3.62	-1.96	-0.49	0.84	1.91
22	.366	.344	-3.52	-1.76	-0.12	1.29	2.72	-3.39	-1.73	-0.27	1.06	2.14
23	.257	.312	-3.86	-2.10	-0.46	0.95	2.38	-3.62	-1.96	-0.50	0.84	1.91
24	.321	.328	-3.46	-1.70	-0.06	1.35	2.78	-3.18	-1.52	-0.06	1.27	2.35
25	.308	.418	-3.77	-2.01	-0.37	1.04	2.47	-3.68	-2.02	-0.56	0.77	1.85
26	.208	.304	-5.14	-3.38	-1.74	-0.33	1.10	-5.09	-3.43	-1.97	-0.64	0.44
27	.308	.346	-3.27	-1.51	0.13	1.54	2.97	-3.40	-1.75	-0.28	1.05	2.12
28	.293	.336	-3.42	-1.66	-0.02	1.39	2.82	-3.28	-1.63	-0.16	1.17	2.24
29	.302	.408	-3.18	-1.42	0.22	1.63	3.06	-2.86	-1.20	0.27	1.60	2.67
30	.276	.312	-3.98	-2.22	-0.58	0.83	2.26	-3.92	-2.26	-0.80	0.54	1.61
ค่าเฉลี่ย	0.30	0.35	-4.01	-2.25	-0.61	0.80	2.23	-3.86	-2.20	-0.73	0.60	1.67

ผลการวิเคราะห์ตามตาราง 13 พบว่าค่าพารามิเตอร์ของข้อสอบของกลุ่มตัวอย่างเพศชายและเพศหญิง เมื่อกลุ่มตัวอย่างมีขนาด 1,000 คน และความยาวของแบบทดสอบขนาด 30 ข้อ กลุ่มตัวอย่างเพศชาย มีค่าอำนาจจำแนกเฉลี่ยเท่ากับ 0.30 และค่าความยากเฉลี่ยในแต่ละลำดับขั้นเท่ากับ -4.01 -2.25 -0.61 0.80 และ 2.23 ส่วนกลุ่มตัวอย่างเพศหญิง มีค่าอำนาจจำแนกเฉลี่ยเท่ากับ 0.35 และค่าความยากเฉลี่ยในแต่ละลำดับขั้นเท่ากับ -3.86 -2.20 -0.73 0.60 และ 1.67

ตาราง 14 ค่าพารามิเตอร์ของข้อสอบ แยกตามกลุ่มเพศชาย – หญิง เมื่อกลุ่มตัวอย่าง 1000 คนและความยาวของแบบทดสอบ 20 ข้อ

ข้อที่	ค่าพารามิเตอร์ของข้อสอบ											
	a _R	a _F	b _{R1}	b _{R2}	b _{R3}	b _{R4}	b _{R5}	b _{F1}	b _{F2}	b _{F3}	b _{F4}	b _{F5}
1	.278	.383	-4.20	-1.38	0.25	1.80	2.97	-3.28	-1.46	0.09	1.45	2.34
2	.327	.399	-5.31	-2.49	-0.87	0.68	1.85	-4.49	-2.68	-1.12	0.23	1.13
3	.216	.363	-5.66	-2.84	-1.22	0.34	1.51	-4.20	-2.39	-0.83	0.53	1.42
4	.201	.216	-6.51	-3.69	-2.07	-0.53	0.65	-6.08	-4.27	-2.72	-1.36	-0.47
5	.276	.308	-4.59	-1.77	-0.15	0.40	2.57	-3.56	-1.75	-0.19	1.17	2.06
6	.224	.294	-4.79	-1.97	-0.35	1.20	2.37	-3.88	-2.07	-0.52	0.84	1.73
7	.316	.408	-4.79	-1.97	-0.35	1.20	2.37	-3.91	-2.10	-0.55	0.81	1.70
8	.283	.249	-6.66	-3.84	-2.22	-0.67	0.51	-5.51	-3.70	-2.15	-0.79	0.10
9	.289	.260	-4.21	-1.39	0.23	0.78	2.95	-3.29	-1.48	0.07	1.43	2.32
10	.252	.332	-5.02	-2.20	-0.58	0.97	2.14	-4.10	-2.29	-0.74	0.62	1.51

ตาราง 14 (ต่อ)

ข้อที่	ค่าพารามิเตอร์ของข้อสอบ											
	a _R	a _F	b _{R1}	b _{R2}	b _{R3}	b _{R4}	b _{R5}	b _{F1}	b _{F2}	b _{F3}	b _{F4}	b _{F5}
11	.264	.309	-4.88	-2.06	-0.44	1.11	2.29	-3.54	-1.73	-0.17	1.18	2.08
12	.249	.291	-5.38	-2.56	-0.94	0.61	1.78	-4.16	-2.34	-0.79	0.57	1.46
13	.207	.308	-4.96	-2.14	-0.52	1.03	2.20	-4.08	-2.27	-0.72	0.64	1.54
14	.301	.341	-4.28	-1.46	0.16	1.71	2.88	-3.39	-1.58	-0.02	1.33	2.23
15	.318	.398	-4.42	-1.60	0.02	1.57	2.75	-3.40	-1.59	-0.04	1.32	2.21
16	.288	.313	-4.65	-1.83	-0.21	1.34	2.51	-3.89	-2.07	-0.52	0.84	1.73
17	.474	.405	-5.32	-2.50	-0.88	0.67	1.84	-4.45	-2.64	-1.08	0.27	1.17
18	.262	.353	-5.52	-2.70	-1.08	0.47	1.64	-4.55	-2.74	-1.19	0.17	1.07
19	.267	.273	-4.82	-1.90	-0.28	1.27	2.44	-3.44	-1.63	-0.08	1.28	2.17
20	.278	.392	-5.11	-2.29	-0.67	0.88	2.05	-3.87	-2.06	-0.51	0.85	1.75
ค่าเฉลี่ย	0.27	0.33	-5.05	-2.23	-0.61	0.94	2.11	-4.05	-2.24	-0.69	0.67	1.56

ผลการวิเคราะห์ตามตาราง 14 พบว่า ค่าพารามิเตอร์ของข้อสอบของกลุ่มตัวอย่างเพศชายและเพศหญิง เมื่อกลุ่มตัวอย่างมีขนาด 1,000 คน และความยาวของแบบทดสอบมีขนาด 20 ข้อ ในกลุ่มตัวอย่างเพศชาย มีค่าอำนาจจำแนกเฉลี่ยเท่ากับ 0.27 และค่าความยากเฉลี่ยในแต่ละลำดับขั้นเท่ากับ -5.05 -2.23 -0.61 0.94 และ 2.11 ส่วนในกลุ่มตัวอย่างเพศหญิง มีค่าอำนาจจำแนกเฉลี่ยเท่ากับ 0.33 และค่าความยากเฉลี่ยในแต่ละลำดับขั้นเท่ากับ -4.05 -2.24 -0.69 0.67 และ 1.56

การที่ค่าพารามิเตอร์ข้อสอบของกลุ่มเพศชายและเพศหญิงตามตาราง 12-14 ทำการประมาณค่ามาจากการวิเคราะห์แยกกันระหว่างกลุ่มผู้สอบย่อย ดังนั้น ก่อนจะนำมาตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจะต้องปรับเทียบสเกลพารามิเตอร์ของข้อสอบให้อยู่บนสเกลเมตริกซ์เดียวกัน ผู้วิจัยได้นำค่าพารามิเตอร์ของข้อสอบของทั้งสองกลุ่มมาคำนวณหา ค่าสัมประสิทธิ์การเชื่อมโยง (Linking Coefficients) คือ ค่าความชันของเส้นถดถอย (A) และค่าจุดตัดบนแกนตั้ง (K) โดยใช้สูตรความสัมพันธ์เชิงเส้น (Linear Relationship) ในการเชื่อมโยง (Linking) (Tate, 1999 : 339) ปรากฏผลดังตาราง 15

ตาราง 15 ค่าสัมประสิทธิ์การเชื่อมโยงของแบบทดสอบขนาด 40 ข้อ, 30 ข้อและ 20 ข้อ

แบบทดสอบขนาด	A	K
40 ข้อ	0.8794	-0.2378
30 ข้อ	0.8601	-0.2392
20 ข้อ	0.8294	-0.1499

ผลการคำนวณค่าสัมประสิทธิ์การเชื่อมโยงตามตาราง 15 พบว่า ค่าความชันของเส้นถดถอย (A) และค่าจุดตัดบนแกนตั้ง (K) ของแบบทดสอบขนาด 40 ข้อ, 30 ข้อและ 20 ข้อ ดังนี้ แบบทดสอบขนาด 40 ข้อ มีค่าความชันของเส้นถดถอย (A) เท่ากับ 0.8794 และค่าจุดตัดบนแกนตั้ง (K) เท่ากับ -0.2378

แบบทดสอบขนาด 30 ข้อ มีค่าความชันของเส้นถดถอย (A) เท่ากับ 0.8601 และค่าจุดตัดบนแกนตั้ง (K) เท่ากับ -0.2392 แบบทดสอบขนาด 20 ข้อ มีค่าความชันของเส้นถดถอย (A) เท่ากับ 0.8294 และค่าจุดตัดบนแกนตั้ง (K) เท่ากับ -0.1499

ค่าสัมประสิทธิ์การเชื่อมโยงที่ได้ตามตาราง 15 ผู้วิจัยนำมาเปรียบเทียบสเกลพารามิเตอร์ของกลุ่มเพศหญิงซึ่งเป็นกลุ่มสนใจ ให้อยู่บนสเกลเมตริกซ์เดียวกันกับสเกลพารามิเตอร์ของกลุ่มเพศชาย ซึ่งเป็นกลุ่มอ้างอิงของการศึกษาครั้งนี้ ปรากฏผลดังตาราง 16 – 18

ตาราง 16 ค่าพารามิเตอร์ของข้อสอบ ที่ปรับเทียบสเกลแล้ว แยกตามกลุ่มเพศชาย-หญิง เมื่อกลุ่มตัวอย่าง 1,000 คนและความยาวของแบบทดสอบ 40 ข้อ

ข้อที่	ค่าพารามิเตอร์ของข้อสอบ											
	a _R	a _F	b _{R1}	b _{R2}	b _{R3}	b _{R4}	b _{R5}	b _{F1}	b _{F2}	b _{F3}	b _{F4}	b _{F5}
1	.314	.290	-3.56	-1.51	0.09	1.58	3.17	-3.22	-1.64	-0.22	0.96	2.09
2	.235	.272	-4.16	-2.11	-0.50	0.98	2.58	-3.55	-1.98	-0.56	0.33	1.75
3	.301	.287	-3.71	-1.66	-0.06	1.42	3.02	-3.21	-1.63	-0.26	0.97	2.09
4	.296	.388	-3.97	-1.92	-0.32	1.17	2.76	-3.72	-2.14	-0.72	0.46	1.58
5	.266	.301	-5.88	-3.83	-2.23	-0.74	0.85	-5.20	-3.63	-2.21	-1.02	0.10
6	.291	.321	-3.58	-1.52	0.08	1.56	3.16	-3.41	-1.84	-0.42	0.77	1.89
7	.246	.310	-3.63	-1.57	0.03	1.51	3.11	-3.19	-1.61	-0.20	0.99	2.11
8	.292	.322	-3.24	-1.19	0.41	1.90	3.49	-2.95	-1.37	0.05	1.23	2.36
9	.239	.323	-4.23	-2.17	-0.57	0.91	2.51	-4.02	-2.44	-1.03	0.16	1.28
10	.271	.234	-1.15	-2.10	-0.50	0.99	2.58	-4.03	-2.48	-1.06	0.13	1.25
11	.287	.358	-4.00	-1.95	-0.35	1.13	2.73	-3.47	-1.89	-0.47	0.71	1.84
12	.273	.312	-4.71	-2.66	-1.06	0.42	2.02	-3.97	-2.39	-0.97	0.21	1.33
13	.211	.273	-4.51	-2.46	-0.86	0.63	2.22	-3.96	-2.39	-0.97	0.22	1.34
14	.278	.340	-4.77	-2.72	-1.11	0.37	1.96	-4.30	-2.72	-1.31	-0.12	1.00
15	.316	.387	-4.58	-2.53	-0.92	0.56	2.15	-3.95	-2.38	-0.96	0.22	1.34
16	.313	.305	-3.72	-1.66	-0.06	1.42	3.02	-3.49	-1.91	-0.49	0.69	1.82
17	.357	.341	-3.98	-1.92	-0.32	1.16	2.76	-3.29	-1.71	-0.29	0.89	2.01
18	.321	.401	-3.74	-1.69	-0.09	1.40	2.99	-3.36	-1.78	-0.36	0.82	1.95
19	.424	.463	-4.65	-2.60	-1.00	0.48	2.08	-4.31	-2.73	-1.31	-0.13	1.00
20	.252	.270	-5.73	-3.68	-2.08	-0.59	1.00	-5.21	-3.63	-2.21	-1.03	0.10
21	.298	.321	-3.98	-1.93	-0.33	1.15	2.75	-3.67	-2.10	-0.68	0.50	1.63
22	.326	.311	-3.55	-1.50	0.10	1.58	3.18	-3.09	-1.51	-0.09	1.09	2.22
23	.410	.391	-4.53	-2.48	-0.88	0.60	2.20	-4.32	-2.75	-1.33	-0.15	0.98
24	.278	.344	-4.73	-2.68	-1.08	0.40	2.00	-4.37	-2.79	-1.37	-0.19	0.93
25	.253	.271	-3.95	-1.90	-0.30	1.19	2.78	-3.40	-1.82	-0.40	0.78	1.90
26	.308	.303	-3.85	-1.79	-0.19	1.29	2.89	-3.44	-1.86	-0.44	0.74	1.87
27	.275	.262	-3.64	-1.58	0.02	1.50	2.10	-3.45	-1.87	-0.45	0.73	1.85
28	.277	.324	-4.26	-2.20	-0.60	0.88	2.48	-3.84	-2.27	-0.85	0.34	1.46
29	.309	.349	-3.57	-1.52	0.08	1.56	3.16	-3.16	-1.58	-0.16	1.02	2.15
30	.293	.368	-4.46	-2.41	-0.81	0.68	2.27	-4.34	-2.76	-1.34	-0.16	0.97

ตาราง 16 (ต่อ)

ข้อที่	ค่าพารามิเตอร์ของข้อสอบ											
	a _R	a _F	b _{R1}	b _{R2}	b _{R3}	b _{R4}	b _{R5}	b _{F1}	b _{F2}	b _{F3}	b _{F4}	b _{F5}
31	.209	.250	-5.29	-3.24	-1.64	-0.16	1.44	-4.33	-2.76	-1.34	-0.15	0.97
32	.317	.358	-4.49	-2.44	-0.84	0.64	2.24	-3.78	-2.21	-0.79	0.40	1.52
33	.260	.379	-5.04	-2.98	-1.38	0.10	1.70	-4.02	-2.44	-1.02	0.16	1.29
34	.320	.307	-4.61	-2.56	-0.96	0.53	2.12	-4.21	-2.64	-1.22	-0.04	1.09
35	.351	.507	-4.52	-2.46	-0.86	0.62	2.22	-4.07	-2.50	-1.08	0.11	1.23
36	.286	.300	-5.71	-3.66	-2.06	-0.57	1.02	-5.30	-3.72	-2.30	-1.12	0.01
37	.236	.339	-5.76	-3.71	-2.11	-0.63	0.97	-5.35	-3.77	-2.35	-1.17	-0.04
38	.299	.302	-3.98	-1.93	-0.33	1.16	2.75	-3.68	-2.10	-0.69	0.50	1.62
39	.270	.368	-4.01	-1.96	-0.36	1.13	2.72	-3.60	-2.03	-0.61	0.58	1.70
40	.338	.334	-3.80	-1.75	-0.15	1.33	2.93	-3.53	-1.95	-0.53	0.65	1.78
ค่าเฉลี่ย	0.29	0.33	-4.31	-2.25	-0.65	0.83	2.43	-3.87	-2.29	-0.87	0.30	1.43

ผลการวิเคราะห์ตามตาราง 16 พบว่า ค่าพารามิเตอร์ของข้อสอบของกลุ่มตัวอย่างเพศชายและเพศหญิง ที่ปรับเทียบสเกลพารามิเตอร์แล้ว เมื่อกลุ่มตัวอย่างมีขนาด 1,000 คน และความยาวของแบบทดสอบมีขนาด 40 ข้อ กลุ่มตัวอย่างเพศชาย มีค่าอำนาจจำแนกเฉลี่ยเท่ากับ 0.29 และค่าความยากเฉลี่ยในแต่ละลำดับขั้นเท่ากับ -4.31 -2.25 -0.65 0.83 และ 2.43 ส่วนกลุ่มตัวอย่างเพศหญิง มีค่าอำนาจจำแนกเฉลี่ยเท่ากับ 0.33 และค่าความยากเฉลี่ยในแต่ละลำดับขั้นเท่ากับ -3.87 -2.29 -0.87 0.30 และ 1.43

ตาราง 17 ค่าพารามิเตอร์ของข้อสอบ ที่ปรับเทียบสเกลแล้ว แยกตามกลุ่มเพศชาย-หญิง เมื่อกลุ่มตัวอย่าง 1,000 คนและความยาวของแบบทดสอบ 30 ข้อ

ข้อที่	ค่าพารามิเตอร์ของข้อสอบ											
	a _R	a _F	b _{R1}	b _{R2}	b _{R3}	b _{R4}	b _{R5}	b _{F1}	b _{F2}	b _{F3}	b _{F4}	b _{F5}
1	.301	.309	-3.71	-1.95	-0.31	1.10	2.53	-3.24	-1.82	-0.56	0.59	1.51
2	.296	.366	-4.39	-2.63	-0.99	0.42	1.85	-3.66	-2.24	-0.98	0.17	1.09
3	.335	.336	-3.44	-1.68	-0.04	1.37	2.80	-3.16	-1.74	-0.48	0.67	1.59
4	.332	.403	-3.46	-1.70	-0.06	1.35	2.78	-3.09	-1.66	-0.40	0.74	1.67
5	.250	.311	-5.40	-3.64	-2.00	-0.59	0.84	-4.71	-3.29	-2.03	-0.88	0.04
6	.312	.356	-3.69	-1.93	-0.29	1.12	2.55	-3.34	-1.91	-0.65	0.49	1.42
7	.329	.366	-3.27	-1.52	0.13	1.53	2.96	-2.77	-1.35	-0.09	1.06	1.98
8	.446	.429	-4.22	-2.46	-0.82	0.59	2.02	-3.89	-2.46	-1.21	-0.06	0.86
9	.292	.302	-4.41	-2.65	-1.01	0.40	1.83	-4.07	-2.64	-1.39	-0.24	0.69
10	.321	.313	-3.56	-1.80	-0.16	1.25	2.68	-3.15	-1.73	-0.47	0.68	1.60
11	.287	.252	-3.36	-1.60	0.04	1.45	2.88	-3.15	-1.73	-0.47	0.68	1.60
12	.296	.369	-3.95	-2.19	-0.55	0.86	2.29	-3.40	-1.98	-0.72	0.43	1.35
13	.333	.374	-3.30	-1.54	0.10	1.51	2.94	-2.85	-1.43	-0.17	0.98	1.90

ตาราง 17 (ต่อ)

ข้อที่	ค่าพารามิเตอร์ของข้อสอบ											
	a _P	a _F	b _{R1}	b _{R2}	b _{R3}	b _{R4}	b _{R5}	b _{F1}	b _{F2}	b _{F3}	b _{F4}	b _{F5}
14	.310	.419	-4.15	-2.19	-0.75	0.66	2.09	-3.92	-2.49	-1.23	-0.09	0.84
15	.219	.290	-4.94	-3.19	-1.54	-0.14	1.30	-4.00	-2.58	-1.32	-0.17	0.75
16	.310	.347	-4.19	-2.43	-0.79	0.62	2.05	-3.42	-1.99	-0.73	0.42	1.34
17	.273	.405	-4.70	-2.94	-1.30	0.11	1.54	-3.71	-2.29	-1.03	0.12	1.04
18	.349	.547	-4.21	-2.45	-0.81	0.60	2.03	-3.74	-2.31	-1.05	0.09	1.02
19	.315	.294	-5.32	-3.56	-1.92	-0.51	0.92	-4.93	-3.50	-2.24	-1.10	-0.18
20	.248	.315	-5.39	-3.63	-1.99	-0.58	0.85	-5.02	-3.59	-2.34	-1.19	-0.27
21	.309	.372	-3.69	-1.93	-0.29	1.12	2.55	-3.35	-1.92	-0.66	0.48	1.41
22	.366	.344	-3.52	-1.76	-0.12	1.29	2.72	-3.16	-1.73	-0.47	0.67	1.60
23	.257	.312	-3.86	-2.10	-0.46	0.95	2.38	-3.35	-1.92	-0.67	0.48	1.40
24	.321	.328	-3.46	-1.70	-0.06	1.35	2.78	-2.97	-1.55	-0.29	0.86	1.78
25	.308	.418	-3.77	-2.01	-0.37	1.04	2.47	-3.40	-1.98	-0.72	0.43	1.35
26	.208	.304	-5.14	-3.38	-1.74	-0.33	1.10	-4.62	-3.19	-0.93	-0.79	0.14
27	.308	.346	-3.27	-1.51	0.13	1.54	2.97	-3.17	-1.74	-0.48	0.66	1.59
28	.293	.336	-3.42	-1.66	-0.02	1.39	2.82	-3.06	-1.64	-0.38	0.77	1.69
29	.302	.408	-3.18	-1.42	0.22	1.63	3.06	-2.70	-1.27	-0.01	1.13	2.06
30	.276	.312	-3.98	-2.22	-0.58	0.83	2.26	-3.61	-2.18	-0.93	0.22	1.15
ค่าเฉลี่ย	0.30	0.35	-4.01	-2.25	-0.61	0.80	2.23	-3.55	-2.13	-0.87	0.28	1.20

ผลการวิเคราะห์ตามตาราง 17 พบว่า ค่าพารามิเตอร์ของข้อสอบของกลุ่มตัวอย่างเพศชาย และ เพศหญิง ที่ทำการปรับเทียบสเกลพารามิเตอร์แล้ว เมื่อกลุ่มตัวอย่างขนาด 1,000 คน และความยาวของแบบทดสอบขนาด 30 ข้อ กลุ่มตัวอย่างเพศชาย มีค่าอำนาจจำแนกเฉลี่ยเท่ากับ 0.30 และค่าความยากเฉลี่ยในแต่ละลำดับขั้นเท่ากับ -4.01 -2.25 -0.61 0.80 และ 2.23 ส่วนกลุ่มตัวอย่างเพศหญิงมีค่าอำนาจจำแนกเฉลี่ย เท่ากับ 0.35 และค่าความยากเฉลี่ยในแต่ละลำดับขั้นเท่ากับ -3.55 -2.13 -0.87 0.28 และ 1.20

ตาราง 18 ค่าพารามิเตอร์ของข้อสอบ ที่ปรับเทียบสเกลแล้ว แยกตามกลุ่มเพศชาย – หญิง เมื่อกลุ่มตัวอย่าง 1,000 คนและความยาวของแบบทดสอบ 20 ข้อ

ข้อที่	ค่าพารามิเตอร์ของข้อสอบ											
	a _R	a _F	b _{R1}	b _{R2}	b _{R3}	b _{R4}	b _{R5}	b _{F1}	b _{F2}	b _{F3}	b _{F4}	b _{F5}
1	.278	.383	-4.20	-1.38	0.25	1.80	2.97	-2.87	-1.36	-0.08	1.05	1.79
2	.327	.399	-5.31	-2.49	-0.87	0.68	1.85	-3.87	-2.37	-1.08	0.04	0.79
3	.216	.363	-5.66	-2.84	-1.22	0.34	1.51	-3.63	-2.13	-0.84	0.29	1.03
4	.201	.216	-6.51	-3.69	-2.07	-0.53	0.65	-5.19	-3.69	-2.40	-1.28	-0.54
5	.276	.308	-4.59	-1.77	-0.15	0.40	2.57	-3.10	-1.60	-0.31	0.82	1.56

ตาราง 18 (ต่อ)

ข้อที่	ค่าพารามิเตอร์ของข้อสอบ											
	a _R	a _F	b _{R1}	b _{R2}	b _{R3}	b _{R4}	b _{R5}	b _{F1}	b _{F2}	b _{F3}	b _{F4}	b _{F5}
6	.224	.294	-4.79	-1.97	-0.35	1.20	2.37	-3.37	-1.87	-0.58	0.55	1.29
7	.316	.408	-4.79	-1.97	-0.35	1.20	2.37	-3.40	-1.89	-0.60	0.52	1.26
8	.283	.249	-6.66	-3.84	-2.22	-0.67	0.51	4.72	-3.22	-1.93	-0.81	-0.07
9	.289	.260	-4.21	-1.39	0.23	0.78	2.95	-2.88	-1.38	-0.09	1.04	1.78
10	.252	.332	-5.02	-2.20	-0.58	0.97	2.14	-3.55	-2.05	-0.76	0.36	1.11
11	.264	.309	-4.88	-2.06	-0.44	1.11	2.29	-3.08	-1.58	-0.29	0.83	1.57
12	.249	.291	-5.38	-2.56	-0.94	0.61	1.78	-3.60	-2.09	-0.81	0.32	1.06
13	.207	.308	-4.96	-2.14	-0.52	1.03	2.20	-3.54	-2.03	-0.75	0.38	1.12
14	.301	.341	-4.28	-1.46	0.16	1.71	2.88	-2.96	-1.46	-0.17	0.96	1.70
15	.318	.398	-4.42	-1.60	0.02	1.57	2.75	-2.97	-1.47	-0.18	0.94	1.69
16	.288	.313	-4.65	-1.83	-0.21	1.34	2.51	-3.37	-1.87	-0.58	0.54	1.29
17	.474	.405	-5.32	-2.50	-0.88	0.67	1.84	-3.84	-2.34	-1.05	0.08	0.82
18	.262	.353	-5.52	-2.70	-1.08	0.47	1.64	-3.93	-2.42	-0.13	-0.01	0.73
19	.267	.273	-4.82	-1.90	-0.28	1.27	2.44	-3.01	-1.50	-0.22	0.91	1.65
20	.278	.392	-5.11	-2.29	-0.67	0.88	2.05	-3.36	-1.86	-0.57	0.56	1.30
ค่าเฉลี่ย	0.27	0.33	-5.05	-2.23	-0.61	0.94	2.11	-3.51	-2.01	-0.72	0.40	1.15

ผลการวิเคราะห์ตามตาราง 18 พบว่า ค่าพารามิเตอร์ของข้อสอบของกลุ่มตัวอย่างเพศชายและเพศหญิง ที่ทำการปรับเทียบสเกลพารามิเตอร์แล้ว เมื่อกลุ่มตัวอย่างมีขนาด 1,000 คน และความยาวของแบบทดสอบมีขนาด 20 ข้อ กลุ่มตัวอย่างเพศชายมีค่าอำนาจจำแนกเฉลี่ย เท่ากับ 0.27 และค่าความยากเฉลี่ยในแต่ละลำดับขั้นเท่ากับ -5.05 -2.23 -0.61 0.94 และ 2.11 ส่วนกลุ่มตัวอย่างเพศหญิงมีค่าอำนาจจำแนกเฉลี่ย เท่ากับ 0.33 และค่าความยากเฉลี่ยในแต่ละลำดับขั้นเท่ากับ -3.51 -2.01 -0.72 0.40 และ 1.15

เมื่อปรับเทียบสเกลพารามิเตอร์ข้อสอบของกลุ่มผู้สอบให้อยู่บนสเกลเดียวกันตามตาราง 16 – 18 แล้ว จึงทำการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ตามวิธี GPCM โดยใช้วิธีการทดสอบไค-สแควร์ของลอร์ด (Lord's χ^2) โดยใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป EXCEL และ MATLAB ช่วยในการวิเคราะห์ปรากฏผลดังตาราง 19 - 21

ตาราง 19 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GPCM ตามแนวคิดของทฤษฎีการตอบข้อสอบด้วยวิธีของลอร์ด (Lord's χ^2) เมื่อกลุ่มตัวอย่าง 1,000 คนและความยาวของแบบทดสอบ 40 ข้อ

ข้อที่	Lord's χ^2	ข้อที่	Lord's χ^2
(1)	6.0	21	6.9
2	5.0	22	4.7
3	4.1	23	12.4

ตาราง 19 (ต่อ)

ข้อที่	Lord's χ^2	ข้อที่	Lord's χ^2
(4)	14.2*	(24)	9.6
5	26.0**	25	3.4
6	4.0	26	4.9
7	7.8	27	6.7
8	6.7	(28)	6.8
9	13.6*	(29)	6.5
10	6.0	30	15.1*
11	8.0	31	4.5
12	4.3	32	4.9
13	7.1	33	15.2*
14	8.0	34	5.8
15	7.7	35	19.0**
16	7.5	36	5.2
(17)	4.1	37	15.3*
18	10.6	38	6.3
19	11.0	(39)	12.7*
20	3.7	(40)	7.6

$$* \chi^2_{.05} = 12.6$$

$$** \chi^2_{.01} = 16.8$$

ข้อสอบที่อยู่ในเครื่องหมายวงเล็บ () หมายถึง ข้อสอบที่ผู้วิจัยสร้างให้แสดงการทำหน้าที่ต่างกัน(DIF) ที่เอื้อต่อกลุ่มเพศชาย ได้แก่ ข้อ 1, 4, 17, 24, 28, 29, 39 และข้อ 40

ผลการตรวจสอบตามตาราง 19 พบว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิธี GPCM ตามแนวคิดของทฤษฎีการตอบข้อสอบด้วยวิธีการทดสอบไค-สแควร์ของลอร์ด(Lord's χ^2) เมื่อกลุ่มตัวอย่าง 1,000 คนและความยาวของแบบทดสอบ 40 ข้อ พบข้อสอบที่ทำหน้าที่ต่างกัน (DIF) จำนวน 8 ข้อ ได้แก่ ข้อสอบที่ทำหน้าที่ต่างกันที่ระดับนัยสำคัญ .05 จำนวน 6 ข้อ คือ ข้อ 4, 9, 30, 33, 37 และข้อ 39 และข้อสอบที่ทำหน้าที่ต่างกันที่ระดับนัยสำคัญ .01 จำนวน 2 ข้อ คือ ข้อ 5 และข้อ 35 ซึ่งข้อสอบทั้ง 8 ข้อนี้สอดคล้องกับการสร้างให้มีการทำหน้าที่ต่างกันเพียง 2 ข้อ คือ ข้อ 4 และข้อ 39 แต่เป็นการทำหน้าที่ต่างกันที่เอื้อต่อเพศหญิง ซึ่งขัดแย้งกับจุดมุ่งหมายของการสร้างข้อสอบที่ตั้งใจสร้างข้อสอบให้มีการทำหน้าที่ต่างกันที่เอื้อต่อเพศชาย

ตาราง 20 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GPCM ตามแนวคิดของทฤษฎีการตอบข้อสอบด้วยวิธีของลอร์ด (Lord's χ^2) เมื่อกลุ่มตัวอย่าง 1,000 คนและความยาวของแบบทดสอบ 30 ข้อ

ข้อที่	Lord's χ^2	ข้อที่	Lord's χ^2
(1)	5.2	16	5.3
2	7.9	17	17.3**
3	8.5	18	24.6**
4	11.0	19	6.7
5	7.2	20	10.4
6	8.7	(21)	10.5
7	6.4	22	8.0
8	11.9	23	7.1
9	6.7	24	5.4
10	6.1	25	15.5*
(11)	8.1	26	14.0*
12	9.0	27	12.4
13	7.5	28	7.9
14	18.5**	29	13.1*
15	8.7	(30)	7.1

$$* \chi^2_{.05} = 12.6$$

$$** \chi^2_{.01} = 16.8$$

ข้อสอบที่อยู่ในเครื่องหมายวงเล็บ () หมายถึง ข้อสอบที่ผู้วิจัยสร้างให้แสดงการทำหน้าที่ต่างกัน(DIF) ที่เอื้อต่อกลุ่มเพศชาย ได้แก่ ข้อ 1, 11, 21 และข้อ 30

ผลการตรวจสอบตามตาราง 20 พบว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิธี GPCM ตามแนวคิดของทฤษฎีการตอบข้อสอบ (IRT) ด้วยวิธีการทดสอบไค-สแควร์ของ ลอร์ด (Lord's χ^2) เมื่อกลุ่มตัวอย่าง 1,000 คนและความยาวของแบบทดสอบ 30 ข้อ พบข้อสอบที่ทำหน้าที่ต่างกัน (DIF) จำนวน 6 ข้อ ได้แก่ ข้อสอบที่ทำหน้าที่ต่างกันที่ระดับนัยสำคัญ .05 จำนวน 3 ข้อ คือ ข้อ 25, 26 และข้อ 29 และข้อสอบที่ทำหน้าที่ต่างกันที่ระดับนัยสำคัญ .01 จำนวน 3 ข้อ คือ ข้อ 14, 17 และข้อ 18 ซึ่งข้อสอบทั้ง 6 ข้อ ที่ตรวจสอบพบนี้ไม่สอดคล้องกับการสร้างให้มีการทำหน้าที่ต่างกันและพบว่าข้อสอบทำหน้าที่ต่างกันอย่างที่เอื้อต่อเพศหญิง ซึ่งขัดแย้งกับจุดมุ่งหมายของการสร้างข้อสอบที่ตั้งใจสร้างข้อสอบให้เอื้อต่อเพศชาย

ตาราง 21 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GPCM ตามแนวคิดของทฤษฎีการตอบข้อสอบด้วยวิธีของลอร์ด (Lord's χ^2) เมื่อกลุ่มตัวอย่าง 1000 คนและความยาวของแบบทดสอบ 20 ข้อ

ข้อที่	Lord's χ^2	ข้อที่	Lord's χ^2
(1)	19.6**	11	10.2
2	14.8*	12	9.3
3	27.9**	13	17.8**
4	6.2	14	12.8*
5	10.0	15	16.2*
6	12.8	16	11.5
7	17.7**	17	15.6*
8	11.6	18	15.2*
9	10.2	19	8.2
10	14.4*	(20)	19.0*

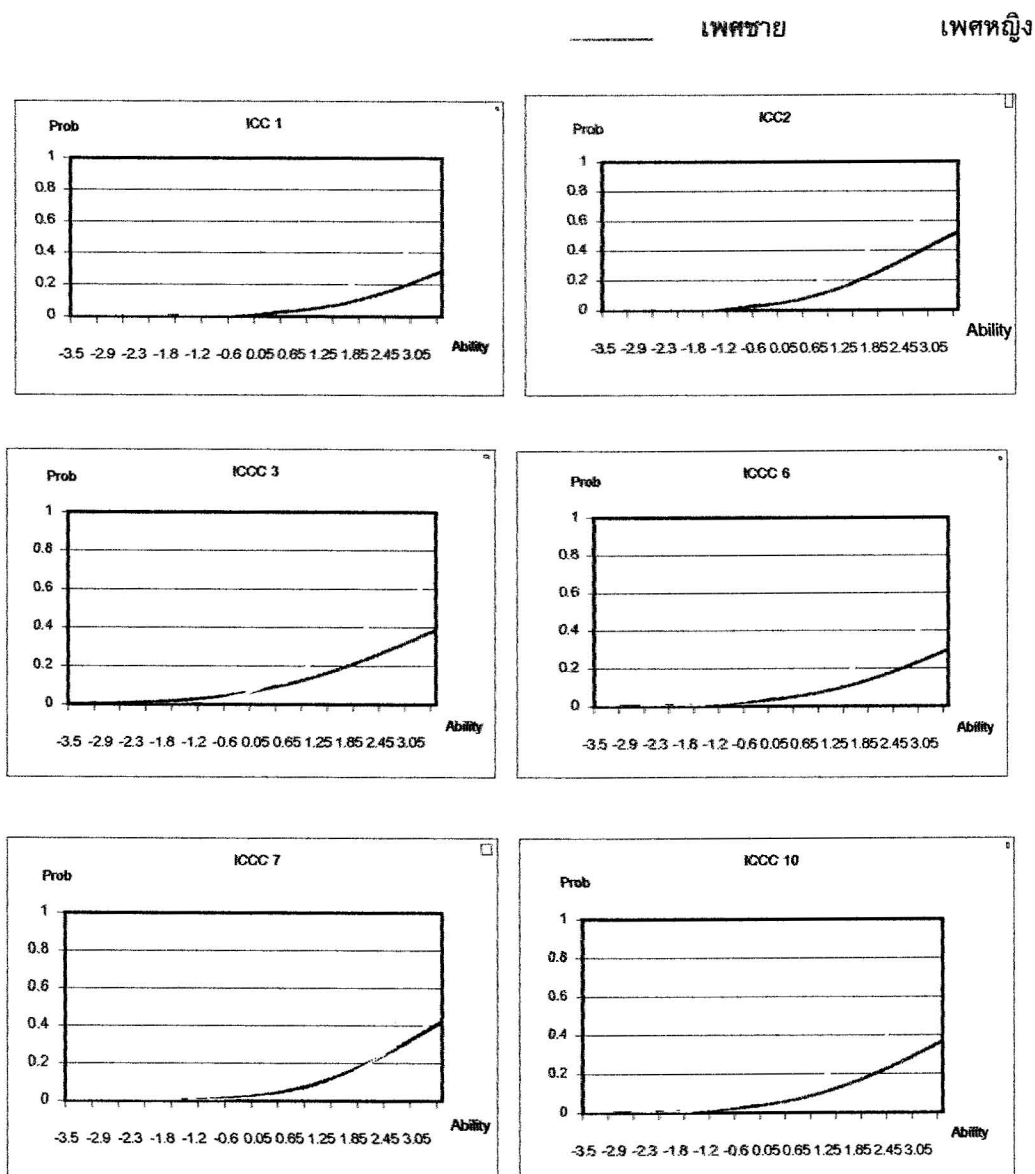
$$* \chi^2_{.05} = 12.6$$

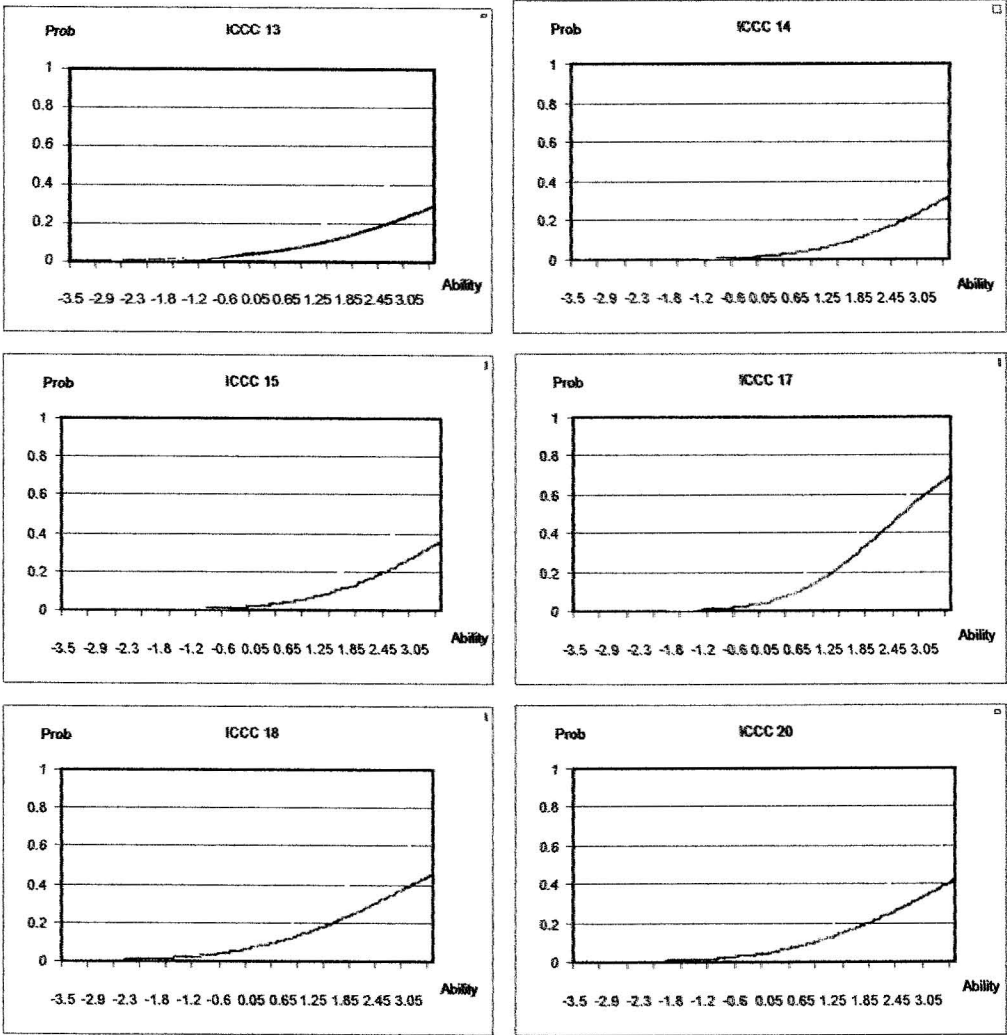
$$** \chi^2_{.01} = 16.8$$

ข้อสอบที่อยู่ในเครื่องหมายวงเล็บ () หมายถึง ข้อสอบที่ผู้วิจัยสร้างให้แสดงการทำหน้าที่ต่างกัน(DIF) ที่เอื้อต่อกลุ่มเพศชาย ได้แก่ ข้อ 1 และข้อ 20

ผลการตรวจสอบตามตาราง 21 พบว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี GPCM ตามแนวคิดของทฤษฎีการตอบข้อสอบ (IRT) ด้วยวิธีการทดสอบไค-สแควร์ของลอร์ด (Lord's χ^2) เมื่อกลุ่มตัวอย่าง 1,000 คนและความยาวของแบบทดสอบ 20 ข้อ พบข้อสอบที่ทำหน้าที่ต่างกัน (DIF) จำนวน 12 ข้อ ได้แก่ ข้อสอบที่ทำหน้าที่ต่างกันที่ระดับนัยสำคัญ .05 จำนวน 7 ข้อ คือ ข้อ 2, 6, 10, 14, 15, 17 และข้อ 18 และข้อสอบที่ทำหน้าที่ต่างกันที่ระดับนัยสำคัญ .01 จำนวน 5 ข้อ คือ ข้อ 1, 3, 7, 13 และข้อ 20 ซึ่งข้อสอบทั้ง 12 ข้อนี้ทำหน้าที่ต่างกันที่เอื้อต่อเพศหญิง ซึ่งขัดแย้งกับจุดมุ่งหมายของการสร้างข้อสอบที่ตั้งใจสร้างข้อสอบให้เอื้อต่อเพศชาย

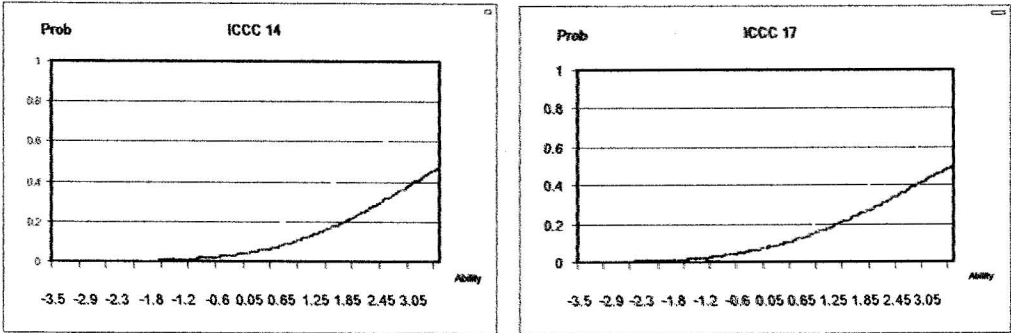
ผลการตรวจสอบตามตาราง 19 - 21 พบว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบตามวิธี GPCM ตามแนวคิดของทฤษฎีการตอบข้อสอบด้วยวิธีการทดสอบไค-สแควร์ของลอร์ด (Lord's χ^2) เมื่อกลุ่มตัวอย่าง 1,000 คนและความยาวของแบบทดสอบ 20 ข้อ, 30 ข้อและ 40 ข้อ พบข้อสอบที่ทำหน้าที่ต่างกัน (DIF) ที่เอื้อต่อเพศหญิง ซึ่งขัดแย้งกับจุดมุ่งหมายของการสร้างข้อสอบที่ตั้งใจสร้างข้อสอบทำหน้าที่ต่างกัน (DIF) ที่เอื้อต่อเพศชาย เพื่อแสดงให้เห็นถึงลักษณะของข้อสอบที่ทำหน้าที่ต่างกัน ผู้วิจัยจึงนำเสนอโค้งคุณลักษณะของข้อสอบ (Item Characteristic Curve: ICC) เฉพาะข้อสอบที่ทำหน้าที่ต่างกันของแบบทดสอบทั้ง 3 ขนาด คือ 20 ข้อ, 30 ข้อและ 40 ข้อ ดังภาพประกอบ 7-9 ส่วนรายละเอียดของโค้งคุณลักษณะของข้อสอบทุกข้อของทั้ง 3 ขนาด ปรากฏในภาคผนวก ฎ

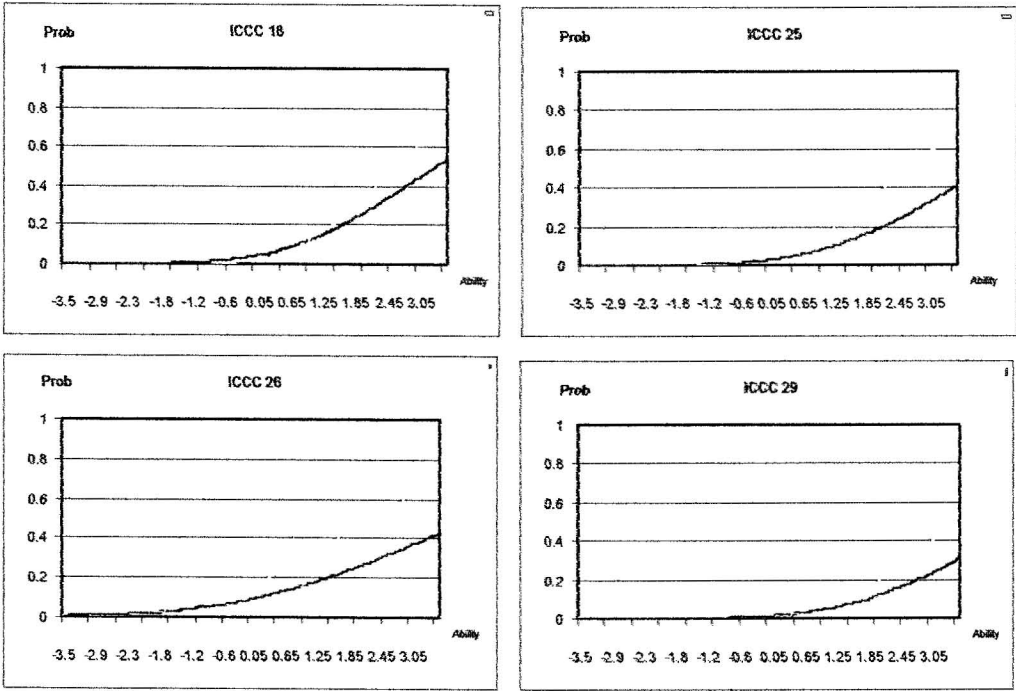




ภาพประกอบ 7 ได้รู้คุณลักษณะข้อสอบ (ICC) ของข้อสอบที่มี DIF ของแบบทดสอบขนาด 20 ข้อ

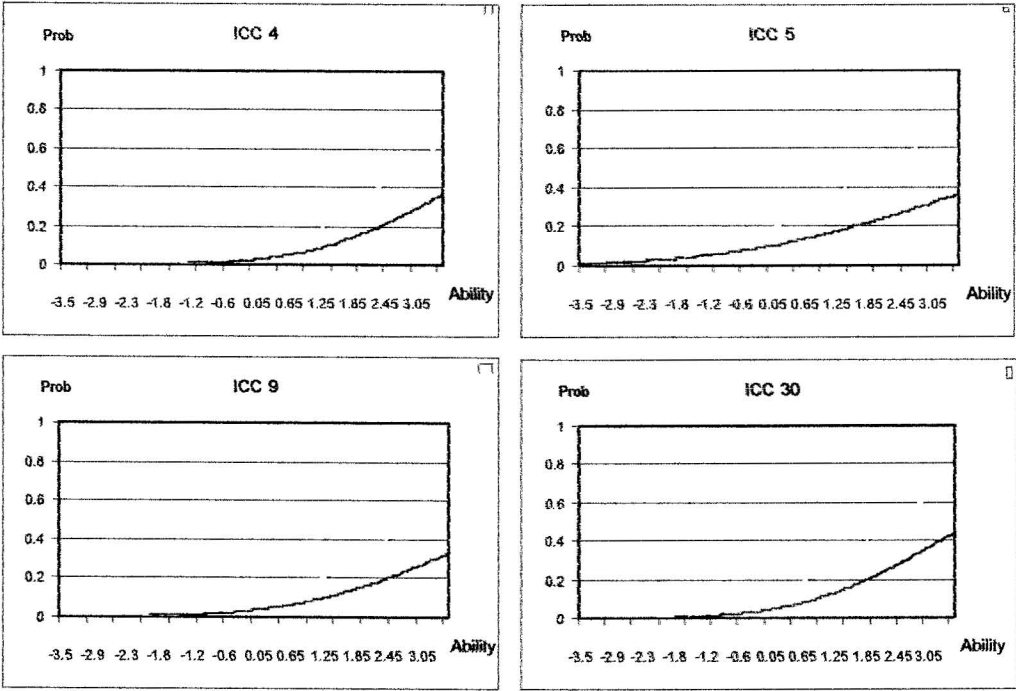
_____ เพศชาย เพศหญิง

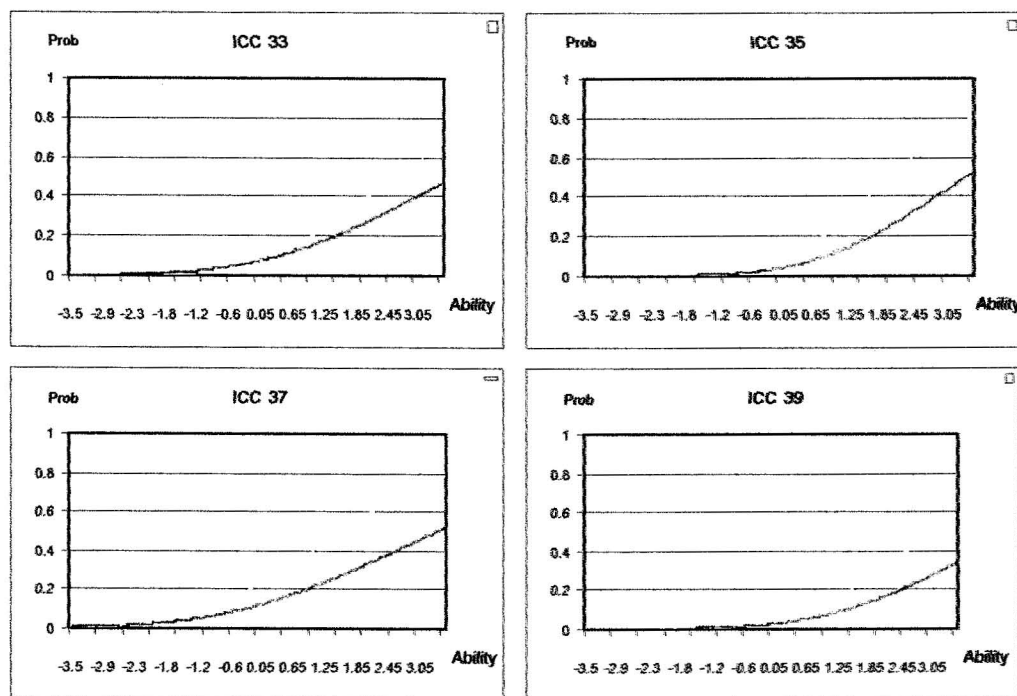




ภาพประกอบ 8 โค้งคุณลักษณะข้อสอบ (ICC) ของข้อสอบที่มี DIF ของแบบทดสอบขนาด 30 ข้อ

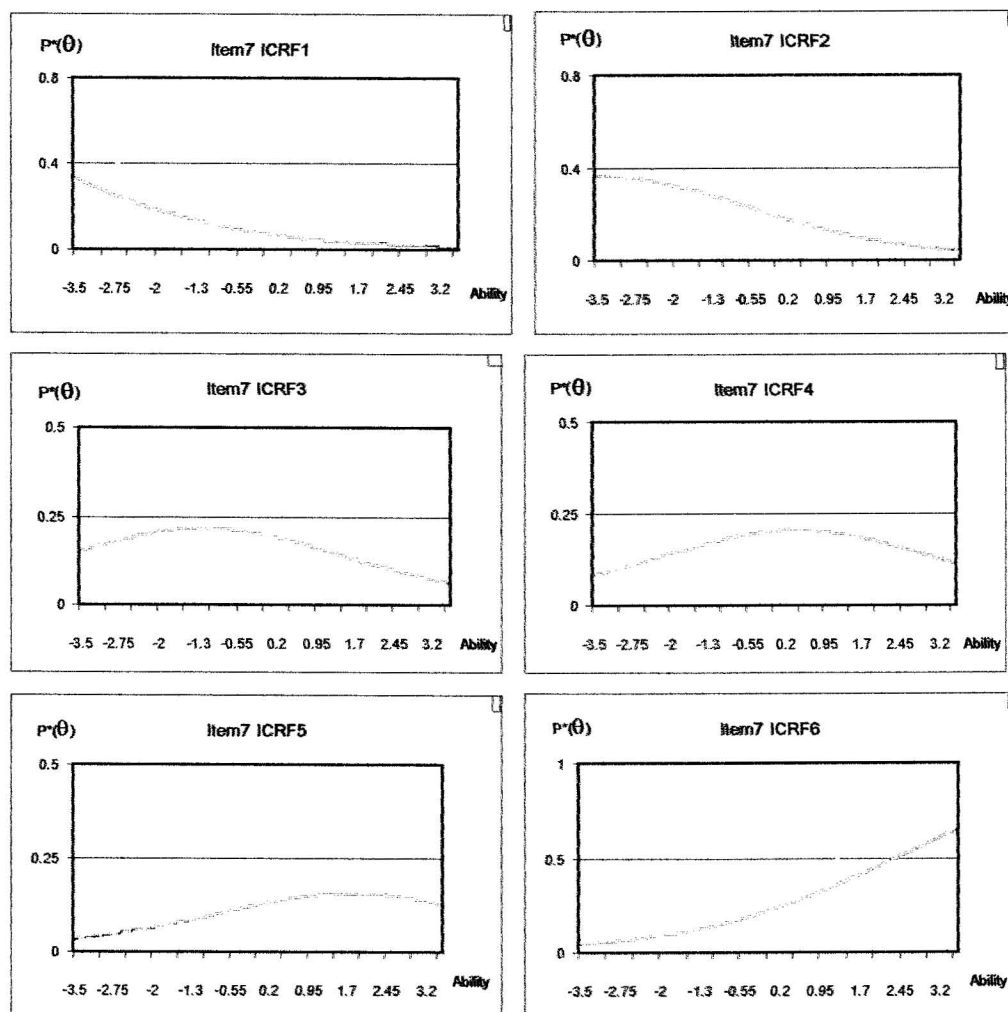
_____ เพศชาย เพศหญิง





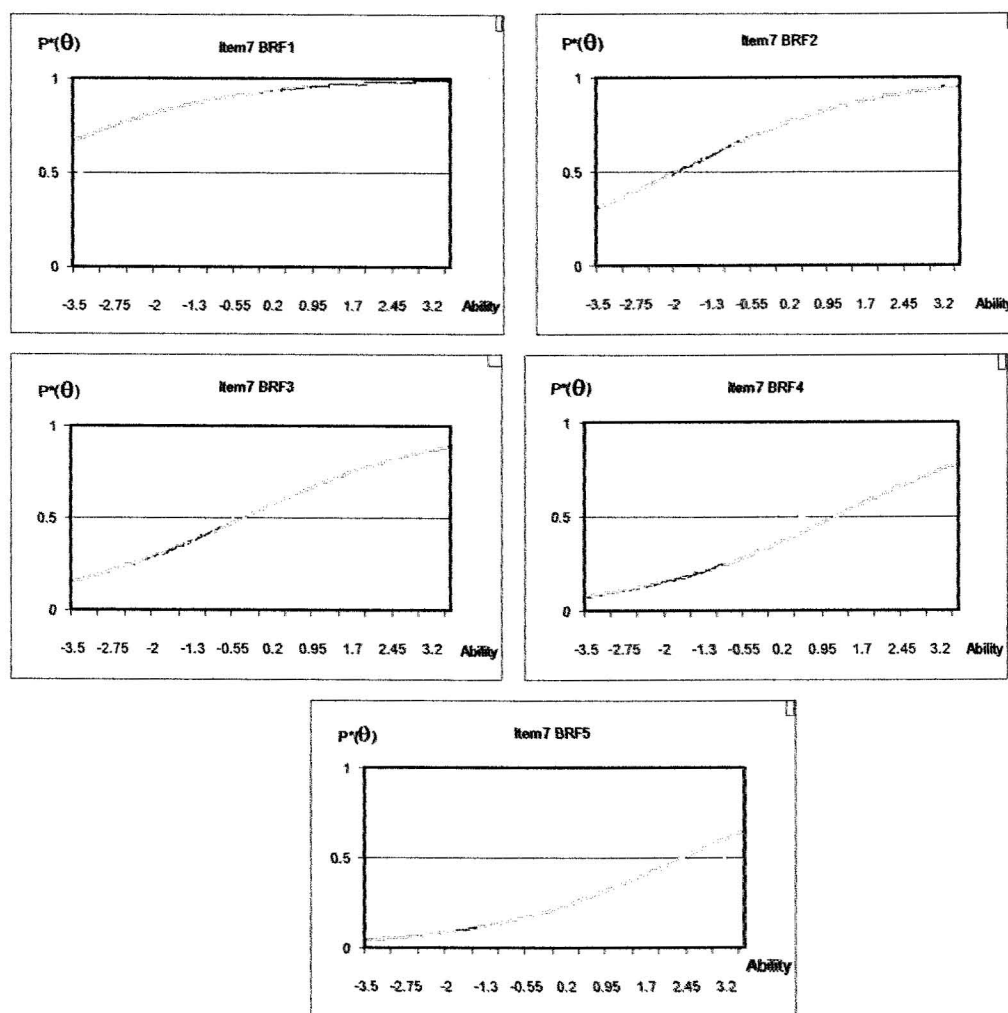
ภาพประกอบ 9 โค้งคุณลักษณะข้อสอบ (ICC) ของข้อสอบที่มี DIF ของแบบทดสอบขนาด 40 ข้อ

โค้งคุณลักษณะข้อสอบ (ICC) มีลักษณะเป็นฟังก์ชันทางคณิตศาสตร์ที่อธิบายความสัมพันธ์ระหว่างระดับความสามารถของผู้สอบ (θ) กับ โอกาส หรือ ความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง $P(\theta)$ ของบุคคลนั้น ซึ่งโค้งคุณลักษณะข้อสอบของข้อสอบแต่ละข้อมีคุณสมบัติไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบ นั่นคือ เมื่อนำข้อสอบไปทดสอบกับผู้สอบ 2 กลุ่มย่อยใดที่มีความสามารถหรือคุณลักษณะที่มุ่งวัดเท่ากัน โค้งคุณลักษณะข้อสอบของผู้สอบ 2 กลุ่มย่อยนั้นจะไม่ต่างกัน ผลการตรวจสอบตามตาราง 19-21 และภาพประกอบ 7-9 แสดงให้เห็น ความต่างกันของโค้งคุณลักษณะข้อสอบของผลการตอบข้อสอบของกลุ่มเพศชายและกลุ่มเพศหญิง นั่นคือ ข้อสอบเกิดการทำหน้าที่ต่างกันระหว่างกลุ่มผู้สอบเพศชายและเพศหญิง เมื่อพิจารณาโค้งคุณลักษณะข้อสอบของข้อสอบทุกข้อที่ตรวจพบว่าเกิดการทำหน้าที่ต่างกันของแบบทดสอบทั้ง 3 ขนาดแล้วพบว่า ข้อสอบเกิดการทำหน้าที่ต่างกันที่เอื้อต่อเพศหญิงแบบสม่ำเสมอ นั่นคือ เพศหญิงมีโอกาสตอบข้อสอบได้ถูกต้องมากกว่าเพศชาย ในทุกระดับความสามารถอย่างชัดเจน ยกเว้นข้อสอบข้อที่ 7 ของแบบทดสอบขนาด 20 ข้อ ที่โค้งคุณลักษณะข้อสอบมีลักษณะที่แตกต่างไปจากข้ออื่น ๆ คือ ข้อสอบเกิดการทำหน้าที่ต่างกันที่เอื้อต่อเพศหญิงแบบไม่สม่ำเสมอ เพื่อแสดงให้เห็นถึงลักษณะของข้อสอบข้อที่ 7 ที่แสดงการทำหน้าที่ต่างกันแบบไม่สม่ำเสมอ ซึ่งต่างไปจากข้อสอบที่แสดงการทำหน้าที่ต่างกันแบบสม่ำเสมอของข้ออื่น ๆ ผู้วิจัยจึงนำเสนอ ฟังก์ชันผลการตอบลำดับขั้นข้อสอบ (Item-Category Response Function : ICRF) ของข้อสอบข้อที่ 7 ของแบบทดสอบขนาด 20 ข้อ ซึ่งการศึกษาครั้งนี้ข้อสอบแต่ละข้อมีลำดับขั้นการตอบ 6 ลำดับขั้น ดังภาพประกอบ 10 ส่วนฟังก์ชันผลการตอบลำดับขั้นข้อสอบ (Item - Category Response Function : ICRF) ของข้อสอบทุกข้อของแบบทดสอบทั้ง 3 ขนาด ปรากฏในภาคผนวก ก



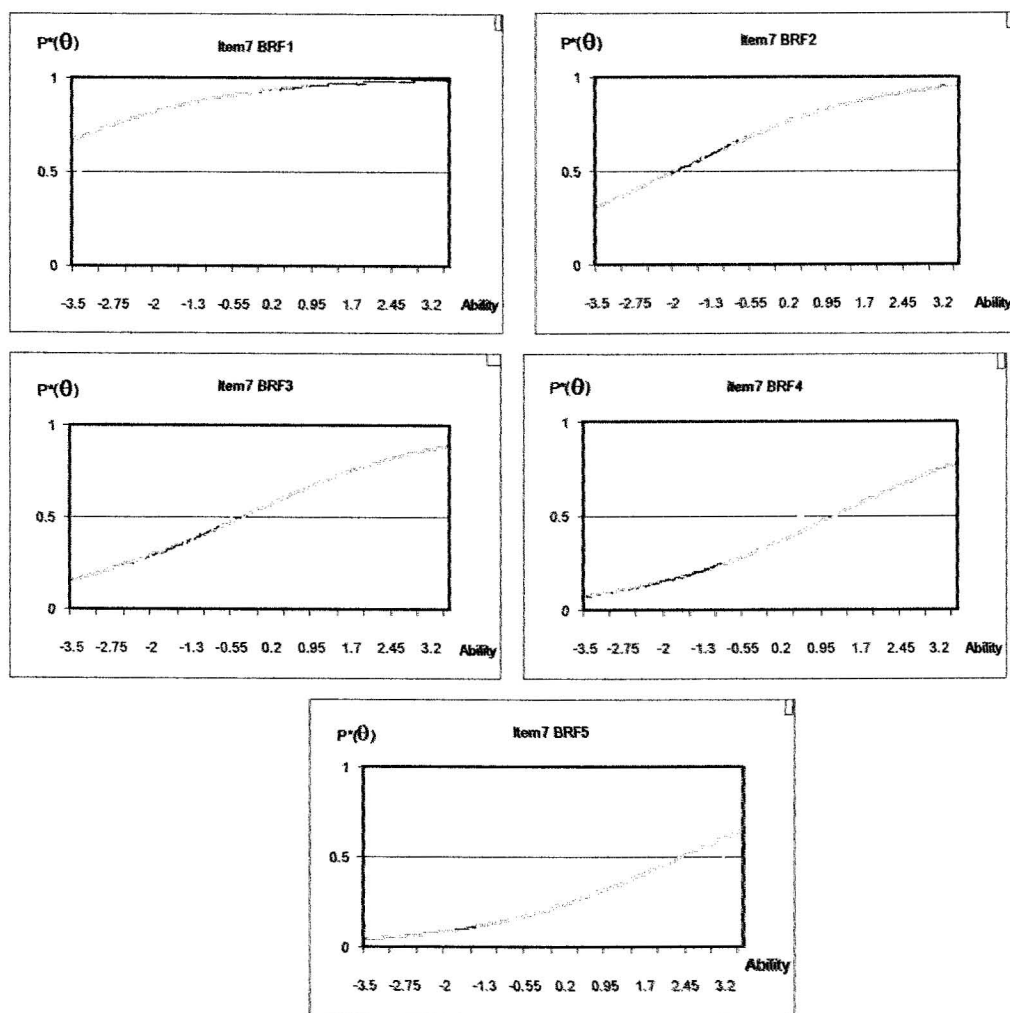
ภาพประกอบ 9 ฟังก์ชันผลการตอบลำดับชั้นข้อสอบ(ICRF)ของข้อสอบข้อ 7 ของแบบทดสอบขนาด 20 ข้อ

ฟังก์ชันผลการตอบลำดับชั้นข้อสอบ (Item - Category Response Function : ICRF) คือ ฟังก์ชันที่แสดงความน่าจะเป็นของการตอบข้อสอบ ในแต่ละลำดับชั้นของข้อสอบ จากลักษณะของฟังก์ชันผลการตอบลำดับชั้นข้อสอบ(ICRF) ของข้อสอบข้อ 7 ของแบบทดสอบขนาด 20 ข้อ เห็นได้ว่า ฟังก์ชันผลการตอบในลำดับชั้นที่ 1 กับลำดับชั้นที่ 6 ของกลุ่มเพศหญิงมีค่าสูงกว่าฟังก์ชันผลการตอบของกลุ่มเพศชาย นั่นคือ ในลำดับชั้นที่ 1 และลำดับชั้นที่ 6 เพศหญิงมีโอกาสตอบข้อสอบได้ถูกต้องมากกว่าเพศชายในทุกระดับความสามารถ แต่ฟังก์ชันผลการตอบในลำดับชั้นที่ 2 - 5 ของกลุ่มเพศชาย มีค่าสูงกว่าฟังก์ชันผลการตอบของกลุ่มเพศหญิง นั่นคือ เพศชายมีโอกาสตอบข้อสอบได้ถูกต้องมากกว่าเพศหญิง ในทุกระดับความสามารถของลำดับชั้นที่ 2 ถึง 5 นอกจากการพิจารณาฟังก์ชันผลการตอบลำดับชั้นข้อสอบ (ICRF) เพิ่มเติมแล้ว เพื่อแสดงให้เห็นถึงลักษณะของข้อสอบข้อที่ 7 ที่แสดงการทำหน้าที่ต่างกันแบบไม่สม่ำเสมอ ผู้วิจัยจึงนำเสนอ ฟังก์ชันผลการตอบข้อสอบในขอบเขตลำดับชั้น (Boundary Response Function : BRF) เพิ่มเติมดังภาพประกอบ 10



ภาพประกอบ 10 ฟังก์ชันผลการตอบข้อสอบในขอบเขตลำดับชั้น (BRF) ของข้อสอบข้อ 7
ของแบบทดสอบขนาด 20 ข้อ

ฟังก์ชันผลการตอบข้อสอบในขอบเขตลำดับชั้น (Boundary Response Function : BRF) คือ ฟังก์ชันที่แสดงความน่าจะเป็นของผลการตอบถูกต้องสะสมในลำดับชั้นที่สูงกว่าลำดับชั้นที่ k หรือ ฟังก์ชันที่แสดงความน่าจะเป็นของผลการตอบถูกต้องสะสมภายในลำดับชั้นที่ $k+1$ ซึ่งการศึกษาครั้งนี้ข้อสอบแต่ละข้อมีลำดับชั้นการตอบ 6 ลำดับชั้น จึงมีความน่าจะเป็นของผลการตอบถูกต้องสะสม (BRF) 5 ลำดับชั้น จากภาพประกอบ 10 พบว่า ฟังก์ชันที่แสดงความน่าจะเป็นของผลการตอบถูกต้องสะสมในขอบเขตลำดับชั้นที่ 2 ถึง 5 มีความไม่สม่ำเสมอในการทำหน้าที่ต่างกัน (Crossing DIF) ส่วนฟังก์ชันผลการตอบในขอบเขตลำดับชั้น (Boundary Response Function : BRF) ของข้อสอบทุกข้อของแบบทดสอบทั้ง 3 ขนาด ปรากฏในภาคผนวก รุ



ภาพประกอบ 10 ฟังก์ชันผลการตอบข้อสอบในขอบเขตลำดับชั้น (BRF) ของข้อสอบข้อ 7
ของแบบทดสอบขนาด 20 ข้อ

ฟังก์ชันผลการตอบข้อสอบในขอบเขตลำดับชั้น (Boundary Response Function : BRF) คือ ฟังก์ชันที่แสดงความน่าจะเป็นของผลการตอบถูกต้องสะสมในลำดับชั้นที่สูงกว่าลำดับชั้นที่ k หรือ ฟังก์ชันที่แสดงความน่าจะเป็นของผลการตอบถูกต้องสะสมภายในลำดับชั้นที่ $k+1$ ซึ่งการศึกษาครั้งนี้ข้อสอบแต่ละข้อมีลำดับชั้นการตอบ 6 ลำดับชั้น จึงมีความน่าจะเป็นของผลการตอบถูกต้องสะสม (BRF) 5 ลำดับชั้น จากภาพประกอบ 10 พบว่า ฟังก์ชันที่แสดงความน่าจะเป็นของผลการตอบถูกต้องสะสมในขอบเขตลำดับชั้นที่ 2 ถึง 5 มีความไม่สม่ำเสมอในการทำหน้าที่ต่างกัน (Crossing DIF) ส่วนฟังก์ชันผลการตอบในขอบเขตลำดับชั้น (Boundary Response Function : BRF) ของข้อสอบทุกข้อของแบบทดสอบทั้ง 3 ขนาด ปรากฏในภาคผนวก รุ

ตอนที่ 3 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วย วิธี GMH และ
วิธี Polytomous SIBTEST

ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วย วิธี GMH และวิธี Polytomous SIBTEST เมื่อกลุ่มตัวอย่างมีขนาด 1,000 , 500 และ 250 คน และแบบทดสอบขนาด 40 , 30 และ 20 ข้อ

เมื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GMH และวิธี Polytomous SIBTEST ในเงื่อนไขต่างๆ แล้ว ผู้วิจัยคำนวณความถูกต้องในการตรวจสอบและความผิดพลาดในการตรวจสอบของวิธีการทั้ง 2 วิธี ในเงื่อนไขขนาดกลุ่มตัวอย่างและความยาวของแบบทดสอบขนาดแตกต่างกัน ปรากฏผลดังรายละเอียดในภาคผนวก ฉ และ สรุปผลดังแสดงในตาราง 22

ตาราง 22 สรุปผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี GMH และวิธี Polytomous SIBTEST ในเงื่อนไขขนาดกลุ่มตัวอย่างและความยาวของแบบทดสอบขนาดแตกต่างกัน

ขนาดกลุ่ม ตัวอย่าง ความยาวแบบทดสอบ		1000				500				250			
		P		E		P		E		P		E	
		$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
40	GMH	57.50	10.54	31.88	4.61	26.25	12.43	16.57	5.11	13.75	14.97	7.82	5.36
ข้อ	Poly-SIB	52.50	11.49	46.25	5.06	28.75	17.73	31.57	6.66	12.50	10.21	21.88	4.89
30	GMH	66.67	7.86	36.25	6.82	38.33	15.81	16.25	9.10	13.34	10.54	10.00	5.62
ข้อ	Poly-SIB	63.34	7.03	56.67	6.57	46.67	13.15	30.00	8.74	31.63	18.34	17.08	6.35
20	GMH	51.67	6.57	72.50	9.86	22.50	8.83	15.83	13.44	10.83	10.43	7.50	8.74
ข้อ	Poly-SIB	69.17	5.62	46.25	8.44	48.34	11.65	35.00	14.19	27.50	18.03	18.75	14.73

ผลการวิเคราะห์ตามตาราง 22 พบว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในเงื่อนไขขนาดกลุ่มตัวอย่างและความยาวของแบบทดสอบขนาดแตกต่างกัน ดังนี้

กรณีที่แบบทดสอบมีขนาด 40 ข้อ เมื่อใช้วิธีการ GMH กับกลุ่มตัวอย่าง 1,000 คน พบว่า ความถูกต้องในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 57.50 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 10.54 ความผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 31.88 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 4.61 เมื่อใช้กับกลุ่มตัวอย่าง 500 คน พบว่า ความถูกต้องในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 26.25 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 12.43 ความผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 16.57 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 5.11 เมื่อใช้กับกลุ่มตัวอย่าง 250 คน พบว่า ความถูกต้องในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 13.75 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 14.97 ความผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 7.82 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 5.36 เมื่อใช้วิธีการ Polytomous SIBTEST กับกลุ่มตัวอย่าง 1,000 คน พบว่า ความถูกต้องในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 52.50 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 11.49 ความผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 46.25 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 5.06 เมื่อใช้กับกลุ่มตัวอย่าง 500 คน พบว่า ความถูกต้องในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 28.75 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 17.73 ความผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 31.57 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 6.66

กรณีที่แบบทดสอบมีขนาด 20 ข้อ เมื่อใช้วิธีการ GMH กับกลุ่มตัวอย่าง 1,000 คน พบว่า ความถูกต้องในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 51.67 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 6.57 ความผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 72.50 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 9.86 เมื่อใช้กับกลุ่มตัวอย่าง 500 คน พบว่า ความถูกต้องในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 22.50 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 8.83 ความผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 15.83 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 13.44 เมื่อใช้กับกลุ่มตัวอย่าง 250 คน พบว่า ความถูกต้องในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 10.83 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 10.43 ความผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 7.50 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 8.74 เมื่อใช้วิธีการ Polytomous SIBTEST กับกลุ่มตัวอย่าง 1,000 คน พบว่า ความถูกต้องในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 69.17 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 5.62 ความผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 46.25 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 8.44 เมื่อใช้กับกลุ่มตัวอย่าง 500 คน พบว่า ความถูกต้องในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 48.34 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 11.65 ความผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 35.00 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 14.19 เมื่อใช้กับกลุ่มตัวอย่าง 250 คน พบว่า ความถูกต้องในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 27.50 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 18.03 ความผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 18.75 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 14.73

การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GMH และวิธี Polytomous SIBTEST

หลังจากตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GMH และวิธี Polytomous SIBTEST แล้ว นำผลการตรวจสอบของแต่ละวิธีมาเปรียบเทียบกัน โดยผู้วิจัยใช้เทคนิคการวิเคราะห์ความแปรปรวนแบบหลายตัวแปร (Multivariate analysis of variance : MANOVA) เพื่อศึกษาผลของกลุ่มตัวอย่างขนาด 1,000, 500 และ 250 คน ที่มีต่อตัวแปรตาม คือ ประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยพิจารณาจากความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (P) และความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (E) พร้อมกัน เมื่อแบบทดสอบมีขนาด 40 , 30 และ 20 ข้อ โดยกำหนดเกณฑ์ในการพิจารณาประสิทธิภาพในการตรวจสอบ ดังนี้ กรณีที่ผลการตรวจสอบมีความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูงและมีความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบต่ำ แสดงถึงประสิทธิภาพในการตรวจสอบสูงสุด ซึ่งเป็นเงื่อนไขที่ต้องการ ส่วนกรณีที่ผลการตรวจสอบมีความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูง แต่มีความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูง หรือ มีความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบต่ำและมีความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบต่ำ หรือมีความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบต่ำ และมีความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูง แสดงถึงประสิทธิภาพในการตรวจสอบต่ำ ซึ่งเป็นเงื่อนไขที่ไม่ต้องการ ดังแสดงในภาพประกอบ 12

		ความถูกต้องในการตรวจสอบ	
		ต่ำ	สูง
ความผิดพลาดในการตรวจสอบ	ต่ำ	ไม่ต้องการ	ต้องการ
	สูง	ไม่ต้องการ	ไม่ต้องการ

ภาพประกอบ 12 เกณฑ์การพิจารณาประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ผู้วิจัยใช้เกณฑ์ดังกล่าวข้างต้นในการพิจารณา โดยนำเสนอผลการเปรียบเทียบดังนี้

1. การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี GMH

การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี GMH เพื่อพิจารณาความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ พร้อมกัน ใช้การวิเคราะห์ความแปรปรวนแบบหลายตัวแปร (MANOVA) ได้ดำเนินการตามลำดับ ดังนี้

1.1 กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ

การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี GMH เมื่อแบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ กลุ่มตัวอย่างขนาด 1,000 , 500 และ 250 คน การทดสอบนัยสำคัญของตัวแปรต้น คือ ขนาดของกลุ่มตัวอย่างที่มีต่อตัวแปรตาม คือ ความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (P) และความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (E) พร้อมกัน พิจารณาจากค่าสถิติ Wilks' Lambda (λ) มีค่า .139 และค่า F เท่ากับ 21.861 พบว่า มีนัยสำคัญทางสถิติที่ระดับ .01 จึงทดสอบนัยสำคัญว่าขนาดของกลุ่มตัวอย่างมีผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (P) และความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (E) เป็นรายตัว พบว่า ขนาดของกลุ่มตัวอย่างมีผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

(P) อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ($F = 31.117$) และขนาดของกลุ่มตัวอย่างมีผลต่อความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (E) อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ($F = 58.42$) ปรากฏผลดังตาราง 23

1.2 กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 30 ข้อ

การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี GMH เมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 ข้อ กลุ่มตัวอย่างขนาด 1,000 , 500 และ 250 คน การทดสอบนัยสำคัญของตัวแปรต้น คือ ขนาดของกลุ่มตัวอย่างที่มีต่อตัวแปรตาม คือ ความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (P) และความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (E) พร้อมกัน พิจารณาจากค่าสถิติ Wilks' Lambda (λ) มีค่า .129 และค่า F เท่ากับ 23.26 พบว่า มีนัยสำคัญทางสถิติที่ระดับ .01 จึงทดสอบนัยสำคัญว่า ขนาดของกลุ่มตัวอย่างมีผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (P) และความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (E) เป็นรายตัว พบว่า ขนาดของกลุ่มตัวอย่างมีผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (P) อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ($F = 50.52$) และ ขนาดของกลุ่มตัวอย่างมีผลต่อความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (E) อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ($F = 35.065$) ปรากฏผลดังตาราง 23

1.3 กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 20 ข้อ

การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี GMH เมื่อแบบทดสอบประกอบด้วยข้อสอบ 20 ข้อ กลุ่มตัวอย่างขนาด 1,000 , 500 และ 250 คน การทดสอบนัยสำคัญของตัวแปรต้น คือ ขนาดของกลุ่มตัวอย่างที่มีต่อตัวแปรตาม คือ ความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (P) และความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (E) พร้อมกัน พิจารณาจากค่าสถิติ Wilks' Lambda (λ) มีค่า .044 และค่า F เท่ากับ 49.324 พบว่า มีนัยสำคัญทางสถิติที่ระดับ .01 จึงทดสอบนัยสำคัญว่า ขนาดของกลุ่มตัวอย่างมีผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (P) และความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (E) เป็นรายตัว พบว่า ขนาดของกลุ่มตัวอย่างมีผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (P) อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ($F = 57.72$) และขนาดของกลุ่มตัวอย่างมีผลต่อความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (E) อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ($F = 105.93$) ปรากฏผลดังตาราง 23

ตาราง 23 ผลการทดสอบนัยสำคัญของขนาดของกลุ่มตัวอย่างที่มีต่อความถูกต้องและความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ พร้อมกันและเป็นรายตัว ด้วยวิธี GMH

ความยาวของ แบบทดสอบ	Wilks' Lambda	F	ค่าสถิติ F	
			P (ความถูกต้อง)	E (ความผิดพลาด)
40	.139	21.861**	31.117**	58.42**
30	.129	23.263**	50.52**	35.065**
20	.044	49.324**	57.722**	105.93**

* มีนัยสำคัญที่ระดับ .05

** มีนัยสำคัญที่ระดับ .01

ผลการทดสอบนัยสำคัญของขนาดของกลุ่มตัวอย่าง คือ 1,000, 500 และ 250 คน ที่มีต่อตัวแปรตาม คือ ความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (P) และความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (E) ปรากฏผลดังตาราง 23 พบว่า ขนาดของกลุ่มตัวอย่างมีผลต่อความถูกต้องและความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ พร้อมกัน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ในทุกขนาดของความยาวของแบบทดสอบ และ เมื่อทดสอบนัยสำคัญของขนาดของกลุ่มตัวอย่างมีผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ และความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เป็นรายตัว พบว่า ขนาดของกลุ่มตัวอย่างมีผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ในทุกขนาดของความยาวของแบบทดสอบ และ ขนาดของกลุ่มตัวอย่างมีผลต่อความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ในทุกขนาดของความยาวของแบบทดสอบ จึงทดสอบความแตกต่างระหว่างค่าเฉลี่ยของความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (P) จากกลุ่มตัวอย่างขนาดต่าง ๆ และทดสอบความแตกต่างระหว่างค่าเฉลี่ยของความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (E) จากกลุ่มตัวอย่างขนาดต่าง ๆ ได้ผลดังตาราง 24

ตาราง 24 ผลการทดสอบความแตกต่างระหว่างค่าเฉลี่ยของความถูกต้องและความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GMH เมื่อแบบทดสอบและกลุ่มตัวอย่างขนาดต่างกัน

ความยาว แบบทดสอบ	ขนาด กลุ่มตัวอย่าง	ความถูกต้องในการตรวจสอบ (P)		
		1,000	500	250
40	1,000	-	31.25**	43.75**
	500	15.312**	-	12.5*
	250	24.062**	8.75**	-
30	1,000	-	28.336**	53.334**
	500	20.00**	-	24.998**
	250	26.25**	6.25	-
20	1,000	-	29.167**	40.833**
	500	56.675**	-	11.666*
	250	65.00**	8.325	-

* มีนัยสำคัญที่ระดับ .05

** มีนัยสำคัญที่ระดับ .01

การเปรียบเทียบความถูกต้องและความผิดพลาดของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GMH เมื่อแบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ, 30 ข้อและ 20 ข้อ กลุ่มตัวอย่างขนาด 1,000, 500 และ 250 คนตามตาราง 24 พบว่า ในแบบทดสอบขนาด 40 ข้อ, 30 ข้อและ 20 ข้อ กลุ่มตัวอย่างขนาด 1,000 คนมีความถูกต้องในการตรวจสอบและความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ แตกต่างกับกลุ่มตัวอย่างขนาด 500 และ 250 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนในแบบทดสอบขนาด 30 ข้อ กลุ่มตัวอย่างขนาด 500 คน มีความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ แตกต่างกับกลุ่มตัวอย่างขนาด 250 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แต่ในแบบทดสอบขนาด 40

ข้อและ 20 ข้อ กลุ่มตัวอย่างขนาด 500 คน กับกลุ่มตัวอย่างขนาด 250 คนมีความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแตกต่างกัน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบขนาด 40 ข้อ กลุ่มตัวอย่างขนาด 500 คน มีความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ แตกต่างกับกลุ่มตัวอย่างขนาด 250 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนแบบทดสอบขนาด 30 ข้อและ 20 ข้อ กลุ่มตัวอย่างขนาด 500 คนกับกลุ่มตัวอย่างขนาด 250 คน มีความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ไม่แตกต่างกัน

2. การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี Polytomous SIBTEST

การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี Polytomous SIBTEST เพื่อพิจารณาความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ พร้อมกัน ใช้การวิเคราะห์ความแปรปรวนแบบหลายตัวแปร (MANOVA) ได้ดำเนินการตามลำดับ ดังนี้

2.1 กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ

การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี Polytomous SIBTEST เมื่อแบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ กลุ่มตัวอย่างขนาด 1,000 , 500 และ 250 คน การทดสอบนัยสำคัญของตัวแปรต้น คือ ขนาดของกลุ่มตัวอย่างที่มีต่อตัวแปรตาม คือ ความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (P) และความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (E) พร้อมกัน พิจารณาจากค่าสถิติ Wilks' Lambda (λ) มีค่า .154 และค่า F เท่ากับ 20.106 พบว่า มีนัยสำคัญทางสถิติที่ระดับ .01 จึงทดสอบนัยสำคัญว่าขนาดของกลุ่มตัวอย่างมีผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (P) และความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (E) เป็นรายตัว พบว่า ขนาดของกลุ่มตัวอย่างมีผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (P) อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ($F = 22.06$) และขนาดของกลุ่มตัวอย่างมีผลต่อความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (E) อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ($F = 48.143$) ปรากฏผลดังตาราง 25

2.2 กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 30 ข้อ

การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี Polytomous SIBTEST เมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 ข้อ กลุ่มตัวอย่างขนาด 1,000 , 500 และ 250 คน การทดสอบนัยสำคัญของตัวแปรต้น คือ ขนาดของกลุ่มตัวอย่างที่มีต่อตัวแปรตาม คือ ความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (P) และความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (E) พร้อมกัน พิจารณาจากค่าสถิติ Wilks' Lambda (λ) มีค่า .119 และค่า F เท่ากับ 24.758 พบว่า มีนัยสำคัญทางสถิติที่ระดับ .01 จึงทดสอบนัยสำคัญว่าขนาดของกลุ่มตัวอย่างมีผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (P) และความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (E) เป็นรายตัว พบว่า ขนาดของกลุ่มตัวอย่างมีผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (P) อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ($F = 13.506$) และขนาดของกลุ่มตัวอย่างมีผลต่อความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (E) อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ($F = 48.143$) ปรากฏผลดังตาราง 25

2.3 กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 20 ข้อ

การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี Polytomous SIBTEST เมื่อแบบทดสอบประกอบด้วยข้อสอบ 20 ข้อ กลุ่มตัวอย่างขนาด 1,000, 500 และ 250 คน การทดสอบนัยสำคัญของตัวแปรต้น คือ ขนาดของกลุ่มตัวอย่างที่มีต่อตัวแปรตาม คือ ความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (P) และความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (E) พร้อมกัน พิจารณาจากค่าสถิติ Wilks' Lambda (λ) มีค่า .199 และค่า F เท่ากับ 16.119 พบว่า มีนัยสำคัญทางสถิติที่ระดับ .01 จึงทดสอบนัยสำคัญว่าขนาดของกลุ่มตัวอย่างมีผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (P) และความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (E) เป็นรายตัว พบว่า ขนาดของกลุ่มตัวอย่างมีผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (P) อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ($F = 26.447$) และขนาดของกลุ่มตัวอย่างมีผลต่อความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (E) อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ($F = 11.713$) ปรากฏผลดังตาราง 25

ตาราง 25 ผลการทดสอบนัยสำคัญของขนาดของกลุ่มตัวอย่างที่มีต่อความถูกต้องและความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ พร้อมกันและเป็นรายตัว ด้วยวิธี Polytomous SIBTEST

ความยาวของ แบบทดสอบ	Wilks' Lambda	F	ค่าสถิติ F	
			P (ความถูกต้อง)	E (ความผิดพลาด)
40	.154	20.106**	22.06**	48.143**
30	.119	24.758**	13.506**	76.432**
20	.199	16.119**	26.447**	11.713**

* มีนัยสำคัญที่ระดับ .05

** มีนัยสำคัญที่ระดับ .01

ผลการทดสอบนัยสำคัญของขนาดของกลุ่มตัวอย่าง คือ 1,000, 500 และ 250 คน ที่มีต่อตัวแปรตาม คือ ความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (P) และความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (E) ปรากฏผลดังตาราง 25 พบว่า ขนาดของกลุ่มตัวอย่างมีผลต่อความถูกต้องและความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ พร้อมกัน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ในทุกขนาดของความยาวของแบบทดสอบ และเมื่อทดสอบนัยสำคัญว่าขนาดของกลุ่มตัวอย่างมีผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ และ ความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เป็นรายตัว พบว่า ขนาดของกลุ่มตัวอย่างมีผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ในทุกขนาดของความยาวของแบบทดสอบ และขนาดของกลุ่มตัวอย่างมีผลต่อความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ในทุกขนาดของความยาวของแบบทดสอบ จึงทดสอบความแตกต่างระหว่างค่าเฉลี่ยของความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (P) จากกลุ่มตัวอย่างขนาดต่าง ๆ และทดสอบความแตกต่างระหว่างค่าเฉลี่ยของความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (E) จากกลุ่มตัวอย่างขนาดต่าง ๆ ได้ผลดังตาราง 26

ตาราง 26 ผลการทดสอบความแตกต่างระหว่างค่าเฉลี่ยของความถูกต้องและความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี Polytomous SIBTEST เมื่อแบบทดสอบและกลุ่มตัวอย่างขนาดต่างกัน

ความยาว แบบทดสอบ	ขนาด กลุ่มตัวอย่าง	ความถูกต้องในการตรวจสอบ (P)		
		1,000	500	250
40	1,000	-	23.75**	40.00**
	500	14.688**	-	16.25*
	250	24.376**	9.688**	-
30	1,000	-	16.67*	31.702**
	500	26.667**	-	15.032*
	250	39.583**	12.919**	-
20	1,000	-	20.833**	41.667**
	500	11.25	-	20.834**
	250	27.5**	16.25*	-

* มีนัยสำคัญที่ระดับ .05

** มีนัยสำคัญที่ระดับ .01

การเปรียบเทียบความถูกต้องและความผิดพลาดของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี Polytomous SIBTEST เมื่อแบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ, 30 ข้อและ 20 ข้อ กลุ่มตัวอย่างขนาด 1,000, 500 และ 250 คน ตามตาราง 26 พบว่า ในแบบทดสอบขนาด 40 ข้อและ 20 ข้อ กลุ่มตัวอย่างขนาด 1,000 คน มีความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ แตกต่างกับกลุ่มตัวอย่างขนาด 500 และ 250 คนอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนในแบบทดสอบขนาด 30 ข้อ กลุ่มตัวอย่างขนาด 1,000 คนมีความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแตกต่างกับกลุ่มตัวอย่างขนาด 500 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และแตกต่างกับกลุ่มตัวอย่างขนาด 250 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบขนาด 40 ข้อ, 30 ข้อและ 20 ข้อ พบว่า ในแบบทดสอบขนาด 40 ข้อและ 30 ข้อ กลุ่มตัวอย่างขนาด 1,000 คน มีความผิดพลาดในการตรวจสอบแตกต่างกับกลุ่มตัวอย่างขนาด 250 คนอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และมีความผิดพลาดในการตรวจสอบแตกต่างกับกลุ่มตัวอย่างขนาด 500 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนในแบบทดสอบขนาด 20 ข้อ กลุ่มตัวอย่างขนาด 1,000 คน กับกลุ่มตัวอย่างขนาด 500 คน ความผิดพลาดในการตรวจสอบ ไม่แตกต่างกัน สำหรับในแบบทดสอบขนาด 20 ข้อ กลุ่มตัวอย่างขนาด 500 คน มีความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแตกต่างกับกลุ่มตัวอย่างขนาด 250 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนในแบบทดสอบขนาด 40 ข้อและ 30 ข้อ กลุ่มตัวอย่างขนาด 500 คนกับกลุ่มตัวอย่างขนาด 250 คนมีความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบขนาด 40 ข้อและ 30 ข้อ กลุ่มตัวอย่างขนาด 500 คนกับกลุ่มตัวอย่างขนาด 250 คน มีความผิดพลาดในการตรวจสอบ แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี GMH และวิธี Polytomous SIBTEST เพื่อพิจารณาความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (P) และความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ(E) พร้อมกัน ใช้การวิเคราะห์ความแปรปรวนแบบหลายตัวแปร (MANOVA) ได้ดำเนินการตามลำดับ ดังนี้

การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี GMH และวิธี Polytomous SIBTEST เมื่อแบบทดสอบประกอบด้วยข้อสอบ 40 ข้อและกลุ่มตัวอย่างขนาด 1,000 คน การทดสอบนัยสำคัญของตัวแปรต้น คือ วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่มีต่อตัวแปรตาม คือ ความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (P) และความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (E) พร้อมกัน พิจารณาจากค่าสถิติ Wilks' Lambda (λ) มีค่า .264 และค่า F เท่ากับ 23.72 พบว่า มีนัยสำคัญทางสถิติที่ระดับ .01 ปรากฏผลดังตาราง 27

การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี GMH และวิธี Polytomous SIBTEST เมื่อแบบทดสอบประกอบด้วยข้อสอบ 40 ข้อและกลุ่มตัวอย่างขนาด 500 คน การทดสอบนัยสำคัญของตัวแปรต้น คือ วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่มีต่อตัวแปรตาม คือ ความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (P) และความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (E) พร้อมกัน พิจารณาจากค่าสถิติ Wilks' Lambda (λ) มีค่า .334 และค่า F เท่ากับ 16.966 พบว่า มีนัยสำคัญทางสถิติที่ระดับ .01 ปรากฏผลดังตาราง 27

การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี GMH และวิธี Polytomous SIBTEST เมื่อแบบทดสอบประกอบด้วยข้อสอบ 40 ข้อและกลุ่มตัวอย่างขนาด 250 คน การทดสอบนัยสำคัญของตัวแปรต้น คือ วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่มีต่อตัวแปรตาม คือ ความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (P) และความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (E) พร้อมกัน พิจารณาจากค่าสถิติ Wilks' Lambda (λ) มีค่า .306 และค่า F เท่ากับ 19.256 พบว่า มีนัยสำคัญทางสถิติที่ระดับ .01 ปรากฏผลดังตาราง 27

การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี GMH และวิธี Polytomous SIBTEST เมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 ข้อและกลุ่มตัวอย่างขนาด 1,000 คน การทดสอบนัยสำคัญของตัวแปรต้น คือ วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่มีต่อตัวแปรตาม คือ ความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (P) และความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (E) พร้อมกัน พิจารณาจากค่าสถิติ Wilks' Lambda (λ) มีค่า .260 และค่า F เท่ากับ 24.208 พบว่า มีนัยสำคัญทางสถิติที่ระดับ .01 ปรากฏผลดังตาราง 27

ตาราง 27 ผลการทดสอบนัยสำคัญของวิธีการตรวจสอบที่มีต่อความถูกต้องและความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ พร้อมกัน

ความยาวของแบบทดสอบ	ขนาดกลุ่มตัวอย่าง	Wilks' Lambda	F
40	1,000	.264	23.720**
	500	.334	16.966**
	250	.306	19.256**
30	1,000	.260	24.208**
	500	.553	6.864*
	250	.531	7.510*
20	1,000	.169	41.727**
	500	.286	21.254**
	250	.517	7.951**

* มีนัยสำคัญที่ระดับ .05

** มีนัยสำคัญที่ระดับ .01

ผลการวิเคราะห์ตามตาราง 27 เป็นการทดสอบนัยสำคัญของตัวแปรต้น คือ วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่มีต่อตัวแปรตาม คือ ความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (P) และความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (E) พร้อมกัน พิจารณาจากค่าสถิติ Wilks' Lambda (λ) พบว่า ในทุกเงื่อนไขขนาดกลุ่มตัวอย่างและความยาวของแบบทดสอบ มีนัยสำคัญทางสถิติ จึงทดสอบความแตกต่างระหว่างค่าเฉลี่ยของความถูกต้องของวิธีการในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (P) และทดสอบความแตกต่างระหว่างค่าเฉลี่ยของความผิดพลาดของวิธีการในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (E) เป็นรายตัว ผลการทดสอบปรากฏดังตาราง 28

ตาราง 28 ผลการทดสอบความแตกต่างระหว่างค่าเฉลี่ยของความถูกต้องในการตรวจสอบและความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ตามวิธี GMH และวิธี Polytomous SIBTEST ในทุกเงื่อนไขขนาดกลุ่มตัวอย่างและความยาวของแบบทดสอบ ที่ต่างกัน

ขนาด กลุ่ม ตัวอย่าง	ความยาวแบบทดสอบ	40 ข้อ	30 ข้อ	20 ข้อ
	ความถูกต้อง (P) ความผิดพลาด (E)	(ค่าเฉลี่ยของวิธี GMH - ค่าเฉลี่ยของวิธี Poly-SIB) ผลต่างของค่าเฉลี่ย		
1000	P	(57.5 - 52.5)	(66.67 - 63.34)	(51.67 - 69.17)
		5.0	3.333	-17.502**
	E	(31.88 - 46.25)	(36.25 - 56.67)	(72.5 - 46.25)
		-14.376**	-20.416**	26.25**
500	P	(26.25 - 28.75)	(38.33 - 46.67)	(22.5 - 48.34)
		-2.5	-8.333	-25.836**
	E	(16.57 - 31.57)	(16.25 - 30.0)	(15.83 - 35.0)
		-15.0**	-13.752*	-19.175*
250	P	(13.75 - 12.5)	(13.34 - 31.63)	(10.83 - 27.5)
		1.25	-18.299*	-16.668*
	E	(7.82 - 21.86)	(10.0 - 17.08)	(7.5 - 18.75)
		-14.062**	-7.083*	-11.25

* มีนัยสำคัญที่ระดับ .05

** มีนัยสำคัญที่ระดับ .01

ผลการวิเคราะห์ตามตาราง 28 การพิจารณาความถูกต้องและความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบตามวิธี GMH และวิธี Polytomous SIBTEST ในเงื่อนไขขนาดกลุ่มตัวอย่างและความยาวของแบบทดสอบแตกต่างกันทั้งหมด 9 เงื่อนไข สรุปผลได้ดังนี้

ในกลุ่มตัวอย่างขนาด 1,000 คน ความถูกต้องในการตรวจสอบวิธี GMH สูงกว่าวิธี Polytomous SIBTEST ในแบบทดสอบขนาด 40 ข้อและ 30 ข้ออย่างไม่มีนัยสำคัญ แต่ในกรณีแบบทดสอบขนาด 20 ข้อวิธี Polytomous SIBTEST สูงกว่าวิธี GMH อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนในแบบทดสอบขนาด 40 ข้อและ 30 ข้อ ความผิดพลาดในการตรวจสอบของวิธี GMH ต่ำกว่าวิธี Polytomous SIBTEST อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แต่ในกรณีแบบทดสอบขนาด 20 ข้อวิธี Polytomous SIBTEST ต่ำกว่าวิธี GMH อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 นั่นคือ ในกรณีแบบทดสอบขนาด 20 ข้อ วิธี Polytomous SIBTEST มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบมากกว่าวิธี GMH อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนในแบบทดสอบขนาด 40 ข้อและ 30 ข้อวิธี GMH มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบมากกว่าวิธี Polytomous SIBTEST อย่างไม่มีนัยสำคัญในด้านความถูกต้องในการตรวจสอบ แต่ด้านความผิดพลาดในการตรวจสอบมีนัยสำคัญทางสถิติที่ระดับ .01

ในกลุ่มตัวอย่างขนาด 500 คน และความยาวของแบบทดสอบขนาด 40 และ 30 ข้อ ความถูกต้องในการตรวจสอบของวิธี Polytomous SIBTEST สูงกว่าวิธี GMH อย่างไม่มีนัยสำคัญ ส่วนความผิดพลาดในการตรวจสอบของวิธี Polytomous SIBTEST สูงกว่าวิธี GMH อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ในแบบทดสอบขนาด 40 ข้อ และอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ในแบบทดสอบขนาด 30 ข้อ ส่วนในแบบทดสอบขนาด 20 ข้อ ความถูกต้องในการตรวจสอบของวิธี Polytomous SIBTEST สูงกว่าวิธี GMH อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนความผิดพลาดในการตรวจสอบของวิธี Polytomous SIBTEST สูงกว่าวิธี GMH อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 นั่นคือ ในกรณีกลุ่มตัวอย่างขนาด 500 คน ประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของวิธีการ 2 วิธี เป็นไปตามเกณฑ์การพิจารณาประสิทธิภาพในเงื่อนไขที่ไม่ต้องการ จึงไม่สามารถระบุได้ชัดเจน วิธีใดมีประสิทธิภาพมากกว่าวิธีใด หรืออาจพิจารณาได้ว่า วิธีการตรวจสอบ 2 วิธี มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ไม่แตกต่างกัน

ในกลุ่มตัวอย่างขนาด 250 คน ของแบบทดสอบขนาด 40 ข้อ ความถูกต้องในการตรวจสอบของวิธี GMH สูงกว่าวิธี Polytomous SIBTEST อย่างไม่มีนัยสำคัญ แต่ความผิดพลาดในการตรวจสอบของวิธี GMH ต่ำกว่าวิธี Polytomous SIBTEST อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนในแบบทดสอบขนาด 30 ข้อความถูกต้องในการตรวจสอบของวิธี Polytomous SIBTEST สูงกว่าวิธี GMH อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่ความผิดพลาดในการตรวจสอบของวิธี Polytomous SIBTEST ก็สูงกว่าวิธี GMH อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ด้วย และในแบบทดสอบขนาด 20 ข้อ ความถูกต้องในการตรวจสอบของวิธี Polytomous SIBTEST สูงกว่าวิธี GMH อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่ความผิดพลาดในการตรวจสอบของวิธี Polytomous SIBTEST ก็สูงกว่าวิธี GMH ด้วยอย่างไม่มีนัยสำคัญ นั่นคือ ในกรณีกลุ่มตัวอย่างขนาด 250 คน ประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของ 2 วิธี เป็นไปตามเกณฑ์การพิจารณาประสิทธิภาพในเงื่อนไขที่ไม่ต้องการ จึงไม่สามารถระบุได้ชัดเจน ว่าวิธีใดมีประสิทธิภาพมากกว่าวิธีใด หรืออาจพิจารณาได้ว่าวิธีการตรวจสอบ 2 วิธีนี้ มีประสิทธิภาพไม่แตกต่างกัน

บทที่ 5

สรุปอภิปรายผล และข้อเสนอแนะ

การศึกษาค้นคว้าครั้งนี้มีจุดมุ่งหมายเพื่อศึกษาวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในแบบทดสอบที่มีการให้คะแนนแบบหลายค่า โดยมีวัตถุประสงค์เฉพาะเพื่อเปรียบเทียบผลการตรวจสอบของวิธี GMH และวิธี Polytomous SIBTEST ในด้านต่าง ๆ ดังนี้

1. เปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีเดียวกัน ในเงื่อนไขขนาดของกลุ่มตัวอย่างและความยาวของแบบทดสอบแตกต่างกัน

2. เปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีต่างกัน ในเงื่อนไขขนาดของกลุ่มตัวอย่างและความยาวของแบบทดสอบแตกต่างกัน

สมมุติฐานในการวิจัยมีดังนี้

1. กลุ่มตัวอย่างขนาดต่างกัน ส่งผลต่อความถูกต้องในการตรวจสอบและความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี GMH มีค่าแตกต่างกัน ในทุกขนาดของความยาวของแบบทดสอบ

2. กลุ่มตัวอย่างขนาดต่างกัน ส่งผลต่อความถูกต้องในการตรวจสอบและความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี Polytomous SIBTEST มีค่าแตกต่างกัน ในทุกขนาดของความยาวของแบบทดสอบ

3. เมื่อพิจารณาความถูกต้องในการตรวจสอบและความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ พร้อมกัน วิธี Polytomous SIBTEST ให้ผลในการตรวจสอบดีกว่าวิธี GMH ในทุกเงื่อนไขขนาดของกลุ่มตัวอย่างและความยาวของแบบทดสอบ

กลุ่มตัวอย่างที่ใช้ในการศึกษาเป็นนักเรียนเพศชายและเพศหญิง ชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1 ปีการศึกษา 2545 ของโรงเรียนสังกัดกรมสามัญศึกษา จังหวัดพระนครศรีอยุธยา ซึ่งเลือกมาโดยการสุ่มแบบสองขั้นตอน (Two-Stage Random Sampling) มีขนาดของโรงเรียนเป็นชั้น (strata) และมีห้องเรียนเป็นหน่วยของการสุ่ม (Sampling Unit) ได้โรงเรียนจำนวน 14 โรงเรียน จำนวน 59 ห้องเรียน นักเรียนทั้งหมดจำนวน 2,247 คน เป็นเพศชาย 1,025 คน เพศหญิง 1,222 คน ใช้นักเรียนทั้งหมดของแต่ละเพศเป็นประชากรเทียม สุ่มกลุ่มตัวอย่างขนาด 1,000 คน, 500 คนและ 250 คน ขนาดละ 10 กลุ่ม เพื่อศึกษาความคงเส้นคงวาในการตรวจสอบการทำหน้าที่ต่างกัน เมื่อกลุ่มตัวอย่างขนาดแตกต่างกันและแบบทดสอบมีขนาดแตกต่างกัน 3 ขนาด คือ 40 ข้อ, 30 ข้อและ 20 ข้อ

เครื่องมือที่ใช้ในการวิจัยครั้งนี้เป็นแบบทดสอบวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 3 เรื่องวงกลม ที่ผู้วิจัยสร้างขึ้นมีลักษณะเป็นข้อสอบปรนัยเลือกตอบแบบถูกผิด (Multiple True-False Item) ชนิด 5 ตัวเลือก โดยให้คะแนนทุกตัวเลือก จำนวน 40 ข้อ มีข้อสอบที่ผู้เชี่ยวชาญพิจารณาว่าเป็นข้อสอบที่แสดงการทำหน้าที่ต่างกันเอื้อต่อเพศชาย จำนวน 8 ข้อ ซึ่งมีค่าความยากระหว่าง 0.50 ถึง 0.80 ค่าอำนาจจำแนก 0.2 ถึง 0.61 คัดเลือกข้อสอบตามสัดส่วนในตารางกำหนดข้อสอบจัดทำเป็นแบบทดสอบประกอบด้วยข้อสอบ 30 ข้อ และ 20 ข้อ

วิธีดำเนินการวิจัย

การดำเนินการทำวิจัยครั้งนี้ ผู้วิจัยได้ดำเนินการตามลำดับขั้นตอน ดังนี้

1. ขอความร่วมมือในการเก็บรวบรวมข้อมูลจากผู้อำนวยการสามัญศึกษาจังหวัดพระนครศรีอยุธยา และเก็บรวบรวมข้อมูลจากนักเรียนในโรงเรียนที่เป็นกลุ่มตัวอย่าง

2. ตรวจให้คะแนนผลการตอบข้อสอบวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 3 เรื่องวงกลม เมื่อแบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ ของนักเรียนที่เป็นกลุ่มตัวอย่างทั้งหมด พิจารณาคะแนนผลการตอบของนักเรียนทั้งสองเพศและแยกตามเพศ แล้วพิจารณาคัดเลือกข้อสอบตามสัดส่วนในตารางกำหนดข้อสอบ จัดทำเป็นแบบทดสอบที่ประกอบด้วยข้อสอบ 30 ข้อและ 20 ข้อ พิจารณาคะแนนผลการตอบของนักเรียนทั้งสองเพศและแยกตามเพศ ของแบบทดสอบที่ประกอบด้วยข้อสอบ 30 ข้อและ 20 ข้อ

3. ใช้การวิเคราะห์องค์ประกอบ (Factor Analysis) ด้วยการวิเคราะห์องค์ประกอบหลัก (Principle Component Analysis) หมุนแกนด้วยวิธีแวนแมกซ์ (Varimax) ตามข้อเสนอของลอร์ด เพื่อตรวจสอบความเป็นมิติเดียวของแบบทดสอบ ซึ่งเป็นข้อตกลงเบื้องต้นของการวิเคราะห์การหาหน้าที่ต่างกันของข้อสอบตามวิธี GPCM ตามแนวคิดของทฤษฎีการตอบข้อสอบ (IRT)

4. สุ่มกลุ่มตัวอย่างแบบไม่ใส่คืนขนาด 1,000 คน จากประชากรเทียมของแต่ละเพศที่ตอบข้อสอบวิชาคณิตศาสตร์ ชั้น มัธยมศึกษาปีที่ 3 เมื่อความยาวของแบบทดสอบมีขนาด 40 ข้อ โดยใช้โปรแกรมสำเร็จรูป SPSS for Window ในการสุ่มกลุ่มตัวอย่าง แล้วทำการประมาณค่าพารามิเตอร์ของข้อสอบตามแนวทฤษฎีการตอบข้อสอบ (IRT) โดยใช้โปรแกรม PARSCALE Version 2.2 ของมูรากิ (Muraki) ซึ่งเป็นโปรแกรมที่ใช้ในการวิเคราะห์ข้อสอบที่มีการตรวจให้คะแนนแบบหลายค่า

5. ปรับเทียบสเกลพารามิเตอร์ข้อสอบของทั้งสองกลุ่มด้วยการแปลงค่าเชิงเส้น โดยใช้สูตรในการปรับเทียบ เพื่อให้สเกลพารามิเตอร์ของข้อสอบของทั้งสองกลุ่มอยู่บนสเกลเมตริกซ์เดียวกัน

6. นำผลการปรับเทียบสเกลพารามิเตอร์ในข้อ 5 มาวิเคราะห์การหาหน้าที่ต่างกันของข้อสอบ (DIF) ตามวิธี GPCM เพื่อใช้เป็นวิธีมาตรฐานในการระบุว่าข้อสอบข้อใดหาหน้าที่ต่างกัน โดยใช้วิธีการทดสอบไค-สแควร์ของลอร์ด (Lord's χ^2) ด้วยโปรแกรมสำเร็จรูป MATLAB ช่วยในการวิเคราะห์

7. เตรียมข้อมูลเพื่อใช้ในการตรวจสอบการหาหน้าที่ต่างกันของข้อสอบด้วยวิธี GMH และวิธี Polytomous SIBTEST โดยสุ่มกลุ่มตัวอย่างแบบไม่ใส่คืนขนาด 1,000 คน, 500 คนและ 250 คน จากประชากรเทียมของแต่ละเพศที่ตอบข้อสอบวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 เมื่อแบบทดสอบมีขนาด 40 ข้อ ซึ่งแต่ละขนาดสุ่ม 10 กลุ่มและในกรณีของแบบทดสอบขนาด 30 ข้อและ 20 ข้อ ดำเนินการสุ่มแบบเดียวกัน ดังนั้น แฟ้มข้อมูลผลการตอบข้อสอบของนักเรียนเพศชายและเพศหญิง ในเงื่อนไขความยาวของแบบทดสอบและขนาดของกลุ่มตัวอย่างแตกต่างกันมีจำนวนทั้งหมด $2 \times 3 \times 3 \times 10 = 180$ แฟ้ม (กลุ่มเพศ $\times$ ขนาดกลุ่มตัวอย่าง $\times$ ขนาดของแบบทดสอบ $\times$ จำนวนครั้งของการสุ่ม) ซึ่งแฟ้มทั้งหมดนี้จะใช้ในการตรวจสอบการหาหน้าที่ต่างกันของข้อสอบด้วยวิธี GMH และวิธี Polytomous SIBTEST

8. นำแฟ้มข้อมูลจากข้อ 7 ของกลุ่มเพศชายและเพศหญิงในเงื่อนไขความยาวของแบบทดสอบและกลุ่มตัวอย่างขนาดเดียวกัน แต่ละคู่มาดำเนินการวิเคราะห์เพื่อตรวจสอบการหาหน้าที่ต่างกันของข้อสอบด้วยวิธี GMH โดยจับคู่นักเรียนเพศชายและเพศหญิงที่มีความสามารถเท่ากัน ซึ่งพิจารณาจากคะแนนรวมของผลการตอบข้อสอบวิชาคณิตศาสตร์ ที่เท่ากัน แล้วนำข้อมูลผลการตอบของนักเรียนที่จับคู่แล้วทั้ง 2 กลุ่ม มาสร้างตารางการันจอร์ $2 \times T \times K$ เมื่อ T เป็นจำนวนลำดับขั้นการตอบ และ k เป็นจำนวนของระดับของ

การแบ่งชั้นตัวแปรการจับคู่ ในแต่ละระดับทั้งหมด k ระดับ เพื่อหาอัตราส่วนแต้มต่อรวม α และใช้สถิติ GMH_{χ^2} ทดสอบนัยสำคัญ โดยใช้โปรแกรมสำเร็จรูป EXCEL และ MATLAB ช่วยในการวิเคราะห์

9. นำแฟ้มข้อมูลจากข้อ 7 ของกลุ่มเพศชายและเพศหญิงในเงื่อนไขความยาวของแบบทดสอบและกลุ่มตัวอย่างขนาดเดียวกัน แต่ละคู่มาดำเนินการวิเคราะห์เพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี Polytomous SIBTEST ด้วยโปรแกรมสำเร็จรูป DIFPACK โปรแกรมย่อย Poly-SIBTEST

10. หลังจากวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GMH และวิธี Polytomous SIBTEST แล้วนำผลการวิเคราะห์ที่ได้มาคำนวณความถูกต้องของการตรวจสอบ และความผิดพลาดของการตรวจสอบ ด้วยการเปรียบเทียบกับผลที่ได้จากการตรวจสอบด้วยวิธี GPCM โดยจะใช้เกณฑ์ว่าข้อสอบที่ถูกตรวจสอบพบว่าทำหน้าที่ต่างกันด้วยวิธี GPCM ถือว่าเป็นข้อที่ทำหน้าที่ต่างกัน ถ้าวิธี GMH และวิธี Polytomous SIBTEST ตรวจสอบพบข้อสอบที่ทำหน้าที่ต่างกันตรงกับข้อสอบที่ถูกตรวจสอบพบด้วยวิธี GPCM ถือว่าวิธีการนั้นมีความถูกต้องในการตรวจสอบ แต่ถ้าการตรวจสอบพบว่าข้อสอบทำหน้าที่ต่างกันไม่ตรงกับข้อสอบที่ตรวจสอบพบด้วยวิธี GPCM ถือว่ามีความผิดพลาดในการตรวจสอบ

11. ทดสอบความแตกต่างของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี GMH และวิธี Polytomous SIBTEST ในด้านความถูกต้องในการตรวจสอบ และความผิดพลาดในการตรวจสอบ พร้อมกัน โดยใช้เทคนิคการวิเคราะห์ความแปรปรวนแบบหลายตัวแปร (Multivariate analysis of variance : MANOVA) โดยใช้โปรแกรมสำเร็จรูป SPSS for Windows

12. เปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี GMH และวิธี Polytomous SIBTEST ใช้เกณฑ์ในการเปรียบเทียบประสิทธิภาพในการทำหน้าที่ต่างกันของข้อสอบคือ วิธีการใดมีความถูกต้องในการตรวจสอบ สูง และมีความผิดพลาดประเภทที่ 1 ต่ำ แสดงว่า วิธีการนั้นมีประสิทธิภาพสูงสุดในการทำหน้าที่ต่างกันของข้อสอบ

สรุปผลการวิจัย

ผลการวิจัยแบ่งเป็น 2 ส่วน คือ ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GMH และวิธี Polytomous SIBTEST และการเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GMH และวิธี Polytomous SIBTEST

1. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GMH และวิธี Polytomous SIBTEST ในเงื่อนไขกลุ่มตัวอย่างและความยาวของแบบทดสอบขนาดต่าง ๆ

1.1 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อแบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ ซึ่งวิธี GPCM ตรวจสอบพบข้อสอบที่แสดงการทำหน้าที่ต่างกัน จำนวน 8 ข้อ ได้ผลดังนี้

วิธี GMH เมื่อใช้กลุ่มตัวอย่างขนาด 1,000 คนและแบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ พบว่า ความถูกต้องในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 57.5 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 10.54 ความผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 31.88 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 4.61 เมื่อใช้กลุ่มตัวอย่างขนาด 500 คน พบว่า ความถูกต้องในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 26.25 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 12.43 ความผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 16.57 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 5.11 เมื่อใช้กลุ่มตัวอย่างขนาด 250 คน พบว่า ความถูกต้องในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 13.75 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 14.97 ความผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 7.82 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 5.36

ตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 10.83 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 10.43 ความผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 7.50 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 8.74

วิธี Polytomous SIBTEST เมื่อใช้กลุ่มตัวอย่างขนาด 1,000 คน และ แบบทดสอบประกอบด้วยข้อสอบ 20 ข้อ พบว่า ความถูกต้องในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 69.17 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 5.62 ความผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 46.25 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 8.44 เมื่อใช้กลุ่มตัวอย่างขนาด 500 คน พบว่า ความถูกต้องในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 48.34 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 11.65 ความผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 35.00 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 14.19 เมื่อใช้กลุ่มตัวอย่างขนาด 250 คน พบว่า ความถูกต้องในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 27.50 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 18.03 ความผิดพลาดในการตรวจสอบมีค่าเฉลี่ยคิดเป็นร้อยละ 18.75 ส่วนเบี่ยงเบนมาตรฐานคิดเป็นร้อยละ 14.73

2. การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GMH และวิธี Polytomous SIBTEST ในเงื่อนไขขนาดของกลุ่มตัวอย่างและความยาวของแบบทดสอบขนาดต่าง ๆ ดังนี้

2.1 การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีเดียวกัน

2.1.1 การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GMH

กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ กลุ่มตัวอย่างขนาด 1,000 คน, 500 คน และ 250 คน พบว่า เมื่อกลุ่มตัวอย่างมีขนาด 1,000 คน ส่งผลให้วิธี GMH มีความถูกต้องในการตรวจสอบสูงกว่ากลุ่มตัวอย่างขนาด 500 คน และ 250 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนกลุ่มตัวอย่างขนาด 500 คน ส่งผลให้วิธี GMH มีความถูกต้องในการตรวจสอบสูงกว่ากลุ่มตัวอย่างขนาด 250 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนความผิดพลาดในการตรวจสอบ พบว่า กลุ่มตัวอย่างทุกขนาด ส่งผลให้วิธี GMH มีความผิดพลาดในการตรวจสอบ แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และเมื่อกลุ่มตัวอย่างมีขนาดใหญ่ขึ้น มีความผิดพลาดเพิ่มมากขึ้น

กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 30 ข้อ กลุ่มตัวอย่างขนาด 1,000 คน, 500 คน และ 250 คน พบว่า กลุ่มตัวอย่างขนาด 1,000 คน ส่งผลให้วิธี GMH มีความถูกต้องในการตรวจสอบสูง และมีความผิดพลาดในการตรวจสอบสูงกว่ากลุ่มตัวอย่างขนาด 500 คน และ 250 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และกลุ่มตัวอย่างขนาด 500 คน ส่งผลให้วิธี GMH มีความถูกต้องในการตรวจสอบสูงกว่ากลุ่มตัวอย่างขนาด 250 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แต่กลุ่มตัวอย่างขนาด 500 คน ส่งผลให้วิธี GMH ความผิดพลาดในการตรวจสอบสูงกว่ากลุ่มตัวอย่างขนาด 250 คน อย่างไม่มีนัยสำคัญ

กรณีที่แบบทดสอบประกอบด้วยข้อสอบ 20 ข้อ กลุ่มตัวอย่างขนาด 1,000 คน, 500 คน และ 250 คน พบว่า กลุ่มตัวอย่างขนาด 1,000 คน ส่งผลให้วิธี GMH มีความถูกต้องในการตรวจสอบสูง และมีความผิดพลาดในการตรวจสอบสูงกว่ากลุ่มตัวอย่างขนาด 500 คน และ 250 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และกลุ่มตัวอย่างขนาด 500 คน ส่งผลให้วิธี GMH มีความถูกต้องในการตรวจสอบสูงกว่ากลุ่มตัวอย่างขนาด 250 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่กลุ่มตัวอย่างขนาด 500 คน ส่งผลให้วิธี GMH ความผิดพลาดในการตรวจสอบสูงกว่ากลุ่มตัวอย่างขนาด 250 คน อย่างไม่มีนัยสำคัญ

กรณีทีกลุ่มตัวอย่างขนาด 500 คน และแบบทดสอบทุกขนาด คือ 40 ข้อ, 30 ข้อและ 20 ข้อ ความถูกต้องในการตรวจสอบของวิธี Polytomous SIBTEST ที่มีค่าสูงกว่าวิธี GMH แต่ความผิดพลาดในการตรวจสอบของวิธี Polytomous SIBTEST ก็มีค่าสูงกว่าวิธี GMH ด้วย

กรณีกลุ่มตัวอย่างขนาด 250 คน และแบบทดสอบขนาด 30 ข้อและ 20 ข้อ แม้ความถูกต้องในการตรวจสอบของวิธี Polytomous SIBTEST ที่มีค่าสูงกว่าวิธี GMH แต่ความผิดพลาดในการตรวจสอบของวิธี Polytomous SIBTEST ก็มีค่าสูงกว่าวิธี GMH ด้วย ส่วนกรณีแบบทดสอบขนาด 40 ข้อของกลุ่มตัวอย่าง 250 คน ความถูกต้องในการตรวจสอบของวิธี GMH มีค่าสูงกว่าวิธี Polytomous SIBTEST อย่างไม่มีนัยสำคัญ แม้ความผิดพลาดในการตรวจสอบของวิธี Polytomous SIBTEST มีค่าสูงกว่าวิธี GMH นั่นคือ ผลการเปรียบเทียบเป็นไปตามเกณฑ์การพิจารณาประสิทธิภาพในเงื่อนไขที่ไม่ต้องการ จึงไม่สามารถระบุได้ชัดเจนว่า วิธีใดมีประสิทธิภาพมากกว่าวิธีใด หรืออาจพิจารณาได้ว่าวิธีการตรวจสอบ 2 วิธีนี้มีประสิทธิภาพไม่แตกต่างกัน

สรุปผลการวิจัยโดยภาพรวม ได้ดังนี้

1. เมื่อแบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ, 30 ข้อและ 20 ข้อ กลุ่มตัวอย่างขนาด 1,000 คน, 500 คนและ 250 คน ส่งผลต่อความถูกต้องในการตรวจสอบและความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GMH แตกต่างกัน ซึ่งเป็นไปตามสมมุติฐานข้อที่ 1 กลุ่มตัวอย่างขนาด 1,000 คน ส่งผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูงกว่ากลุ่มตัวอย่างขนาด 500 คน และ 250 คน และความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ก็สูงกว่าด้วย ส่วนกลุ่มตัวอย่างขนาด 500 คน ส่งผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูงกว่ากลุ่มตัวอย่างขนาด 250 คน แต่ความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ไม่แตกต่างกัน
2. เมื่อแบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ, 30 ข้อและ 20 ข้อ กลุ่มตัวอย่างขนาด 1,000 คน, 500 คนและ 250 คน ส่งผลต่อความถูกต้องในการตรวจสอบและความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี Polytomous SIBTEST แตกต่างกัน ซึ่งเป็นไปตามสมมุติฐานข้อที่ 2 กลุ่มตัวอย่างขนาด 1,000 คน ส่งผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูงกว่ากลุ่มตัวอย่างขนาด 500 คนและ 250 คน และกลุ่มตัวอย่างขนาด 500 คน ส่งผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูงกว่ากลุ่มตัวอย่างขนาด 250 คน ส่วนกลุ่มตัวอย่างขนาด 500 คนส่งผลต่อความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สูงกว่า กลุ่มตัวอย่างขนาด 250 คนด้วย แต่กลุ่มตัวอย่างขนาด 1,000 คนและกลุ่มตัวอย่างขนาด 500 คน ส่งผลต่อความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ไม่แตกต่างกัน
3. เมื่อแบบทดสอบประกอบด้วยข้อสอบ 20 ข้อ กลุ่มตัวอย่างขนาด 1,000 คน, 500 คนและ 250 คน ส่งผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี Polytomous SIBTEST สูงกว่าวิธี GMH แต่ส่งผลต่อความผิดพลาดในการตรวจสอบในบางเงื่อนไขขนาดของกลุ่มตัวอย่าง เมื่อแบบทดสอบประกอบด้วยข้อสอบ 40 ข้อและ 30 ข้อ กลุ่มตัวอย่างขนาด 1,000 คน, 500 คน และ 250 คน ส่งผลต่อความถูกต้องและความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ไม่แตกต่างกัน ซึ่งไม่เป็นไปตามสมมุติฐานข้อที่ 3

อภิปรายผลการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์ เพื่อศึกษาวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่เหมาะสมกับข้อมูลการตอบที่มีการให้คะแนนแบบหลายค่า และมีวัตถุประสงค์เฉพาะเพื่อเปรียบเทียบประสิทธิภาพของผลการตรวจสอบของวิธีการตรวจสอบ 2 วิธี คือ วิธี GMH และวิธี Polytomous SIBTEST ไม่ได้มุ่งนำผลการตรวจสอบไปใช้ในการปรับปรุงคุณภาพของแบบทดสอบที่สร้างขึ้น และจากการศึกษาค้นคว้ายังไม่พบว่ามีการศึกษาเพื่อเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจากข้อมูลผลการตอบข้อสอบที่ต้องตอบทุกตัวเลือกและมีการให้คะแนนแบบหลายค่า ด้วยวิธี GMH และวิธี Polytomous SIBTEST แต่มีการเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ จากข้อมูลผลการตอบข้อสอบที่มีการให้คะแนนแบบ 2 ค่า (0 และ 1) ดังนั้น การอ้างอิงส่วนใหญ่จึงเปรียบเทียบกับผลการศึกษาที่เป็นข้อมูลผลการตอบข้อสอบที่มีการให้คะแนนแบบสองค่า ผลการวิจัย พบประเด็นสำคัญสรุปได้ดังนี้

1. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีเดียวกัน

1.1 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GMH

ความถูกต้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GMH ระหว่างกลุ่มตัวอย่างขนาดต่างกัน คือ 1,000 คน, 500 คนและ 250 คน พบว่า ขนาดกลุ่มตัวอย่าง ส่งผลต่อความถูกต้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GMH แตกต่างกัน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 เกือบทุกเงื่อนไข ยกเว้นเมื่อกลุ่มตัวอย่างมีขนาด 500 คนและ 250 คนเมื่อแบบทดสอบมีขนาด 40 ข้อและ 20 ข้อ ที่ส่งผลต่อความถูกต้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GMH แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 นั่นคือ เมื่อกลุ่มตัวอย่างมีขนาดเพิ่มมากขึ้น ความถูกต้องในการตรวจสอบของวิธี GMH มีค่าเพิ่มมากขึ้นด้วย ซึ่งสอดคล้องกับผลการศึกษาของสวามินาทานและโรเจอร์ (Swaminathan & Rogers, 1990) ที่ศึกษาวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี LR กับวิธี MH โดยใช้กลุ่มตัวอย่างขนาด 250 คนและ 500 คน พบว่าวิธี MH ตรวจพบการทำหน้าที่ต่างกันของข้อสอบได้ดีกว่าวิธี LR และมีความถูกต้องในการตรวจสอบร้อยละ 75 เมื่อใช้กลุ่มตัวอย่าง 250 คน และตรวจพบได้ถูกต้องร้อยละ 100 เมื่อใช้กลุ่มตัวอย่างขนาด 500 คนและสอดคล้องกับการศึกษาของจิตินา วรณศรี (2539) ที่เปรียบเทียบประสิทธิภาพของวิธีการ MH และวิธี SIBTEST พบว่า เมื่อกลุ่มตัวอย่างมีขนาด 200 คนและ 600 คน ทั้งสองวิธีสามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ถูกต้องร้อยละ 50 แต่ถ้ากลุ่มตัวอย่างขนาด 1,000 คน สามารถตรวจสอบได้ถูกต้องร้อยละ 100 จากการศึกษาที่นำมาอ้างอิงเป็นการศึกษากับข้อสอบที่มีการให้คะแนนแบบสองค่า แต่การศึกษาดังนี้เป็นการศึกษาเกี่ยวกับข้อสอบที่มีการให้คะแนนแบบหลายค่า พบว่า กรณีที่แบบทดสอบขนาด 40 ข้อ เมื่อใช้วิธี GMH กับกลุ่มตัวอย่างขนาด 250 คน มีความถูกต้องในการตรวจสอบเพียงร้อยละ 14 (13.75%) เมื่อกลุ่มตัวอย่างขนาด 500 คน มีความถูกต้องในการตรวจสอบเพียงร้อยละ 26 (26.25%) นั่นคือ มีความถูกต้องเพิ่มขึ้นร้อยละ 13 (12.5%) แต่เมื่อกลุ่มตัวอย่างขนาด 1,000 คน มีความถูกต้องในการตรวจสอบร้อยละ 58 (57.5%) นั่นคือ มีความถูกต้องเพิ่มขึ้นร้อยละ 31 (31.25%) เมื่อเทียบกับกลุ่มตัวอย่างขนาด 500 คน ส่วนกรณีที่แบบทดสอบขนาด 30 ข้อ เมื่อใช้วิธี GMH กับกลุ่มตัวอย่างขนาด 250 คน มีความถูกต้องในการตรวจสอบเพียงร้อยละ 13 (13.34%) เมื่อกลุ่มตัวอย่างขนาด 500 คนมีความถูกต้องในการตรวจสอบเพียงร้อยละ 38 (38.33%) นั่นคือมีความถูกต้องเพิ่มขึ้นร้อยละ 25 (24.99%) แต่เมื่อกลุ่มตัวอย่างขนาด 1,000 คน มีความถูกต้องในการตรวจสอบร้อยละ 67 (66.67%) นั่นคือ มีความถูกต้องเพิ่มขึ้นร้อยละ 28 (28.34%) เมื่อเทียบกับกลุ่มตัวอย่างขนาด 500 คน กรณีที่แบบทดสอบขนาด 20 ข้อ เมื่อใช้วิธี GMH กับกลุ่มตัวอย่างขนาด 250 คน มีความถูกต้องในการ

ตรวจสอบเพียงร้อยละ 11 (10.83%) เมื่อกลุ่มตัวอย่างขนาด 500 คน มีความถูกต้องในการตรวจสอบเพียงร้อยละ 23 (22.5%) นั่นคือ มีความถูกต้องเพิ่มขึ้นร้อยละ 12 (11.67%) แต่เมื่อกลุ่มตัวอย่างขนาด 1,000 คน มีความถูกต้องในการตรวจสอบร้อยละ 52 (51.67%) นั่นคือ มีความถูกต้องเพิ่มขึ้นร้อยละ 29 (29.17%) เมื่อเทียบกับกลุ่มตัวอย่างขนาด 500 คน

ความผิดพลาดของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GMH ระหว่างกลุ่มตัวอย่างขนาดต่างกัน คือ 1,000 คน, 500 คนและ 250 คน พบว่า ขนาดกลุ่มตัวอย่างส่งผลต่อความผิดพลาดของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GMH แตกต่างกัน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 เกือบทุกเงื่อนไข ยกเว้นเมื่อกลุ่มตัวอย่างมีขนาด 500 คนและ 250 คน เมื่อแบบทดสอบมีขนาด 30 ข้อและ 20 ข้อที่ส่งผลต่อความผิดพลาดของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GMH แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 นั่นคือ เมื่อกลุ่มตัวอย่างมีขนาดเพิ่มขึ้น ความผิดพลาดในการตรวจสอบของวิธี GMH มีค่าเพิ่มมากขึ้นด้วย ซึ่งไม่สอดคล้องกับผลการศึกษาของจิตมา วรณศรี (2539) ที่พบว่า เมื่อกลุ่มตัวอย่างมีขนาด 200 คน วิธี MH มีความผิดพลาดในการตรวจสอบอยู่ในระดับหนึ่ง (อยู่ในช่วง 1% - 9%) เมื่อกลุ่มตัวอย่างเพิ่มเป็น 600 คน ความผิดพลาดในการตรวจสอบเพิ่มขึ้น (อยู่ในช่วง 5% - 9%) แต่เมื่อเพิ่มขนาดกลุ่มตัวอย่างเป็น 1,000 คน ความผิดพลาดในการตรวจสอบลดลงเหลือร้อยละ 0 จากการศึกษาที่นำมาอ้างอิงเป็นการศึกษาข้อสอบที่มีการให้คะแนนแบบสองค่า แต่การศึกษารังนี้เป็นการศึกษากับข้อสอบที่มีการให้คะแนนแบบหลายค่า พบว่า กรณีที่แบบทดสอบขนาด 40 ข้อ เมื่อใช้วิธี GMH กับกลุ่มตัวอย่างขนาด 250 คน มีความผิดพลาดในการตรวจสอบร้อยละ 8 (7.82%) เมื่อกลุ่มตัวอย่างขนาด 500 คน มีความผิดพลาดในการตรวจสอบร้อยละ 17 (16.57%) นั่นคือ มีความผิดพลาดเพิ่มขึ้นร้อยละ 9 (8.85%) แต่เมื่อกลุ่มตัวอย่างขนาด 1,000 คนมีความผิดพลาดในการตรวจสอบร้อยละ 32 (31.88%) นั่นคือ มีความผิดพลาดเพิ่มขึ้นร้อยละ 15 (15.21%) เมื่อเทียบกับกลุ่มตัวอย่างขนาด 500 คน กรณีที่แบบทดสอบขนาด 30 ข้อ เมื่อใช้วิธี GMH กับกลุ่มตัวอย่างขนาด 250 คน มีความผิดพลาดในการตรวจสอบร้อยละ 10 เมื่อกลุ่มตัวอย่างขนาด 500 คน มีความผิดพลาดในการตรวจสอบร้อยละ 16 (16.25%) นั่นคือ มีความผิดพลาดเพิ่มขึ้นร้อยละ 6 (6.25%) แต่เมื่อกลุ่มตัวอย่างขนาด 1,000 คน มีความผิดพลาดในการตรวจสอบร้อยละ 36 (36.25%) นั่นคือ มีความผิดพลาดเพิ่มขึ้นร้อยละ 20 เมื่อเทียบกับกลุ่มตัวอย่างขนาด 500 คน กรณีที่แบบทดสอบขนาด 20 ข้อ เมื่อใช้วิธี GMH กับกลุ่มตัวอย่างขนาด 250 คน มีความผิดพลาดในการตรวจสอบร้อยละ 8 (7.5%) เมื่อกลุ่มตัวอย่างขนาด 500 คน มีความผิดพลาดในการตรวจสอบร้อยละ 16 (15.83%) นั่นคือ มีความผิดพลาดเพิ่มขึ้นร้อยละ 8 (8.33%) แต่เมื่อกลุ่มตัวอย่างขนาด 1,000 คน มีความผิดพลาดในการตรวจสอบร้อยละ 73 (72.5%) นั่นคือ มีความผิดพลาดเพิ่มขึ้นร้อยละ 57 (56.67%) เมื่อเทียบกับกลุ่มตัวอย่างขนาด 500 คน

สรุปผลการตรวจสอบในด้านความผิดพลาดในการตรวจสอบ พบว่า การใช้วิธี GMH ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่า มีความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันเพิ่มขึ้น เมื่อกลุ่มตัวอย่างมีขนาดเพิ่มขึ้น โดยเฉพาะเมื่อกลุ่มตัวอย่างมีขนาด 1,000 คน และแบบทดสอบมีขนาด 20 ข้อ ซึ่งขัดแย้งกับการศึกษาของจิตมา วรณศรี (2539) ดังกล่าวข้างต้น ที่พบว่า ในแบบทดสอบทุกขนาด เมื่อกลุ่มตัวอย่างเป็น 1,000 คน ไม่พบความผิดพลาดในการตรวจสอบ แต่สอดคล้องกับการศึกษาของ รูสโซและสแตต (Roussos & Stout, 1996 : 215-230) ศึกษาผลของการใช้กลุ่มตัวอย่างขนาดเล็กที่มีต่อความผิดพลาดชนิดที่ 1 ด้วยวิธี SIBTEST และวิธี MH โดยใช้กลุ่มตัวอย่างขนาด 100 200 500 และ 1,000 คน ในการศึกษาครั้งที่ 1 และ ใช้กลุ่มตัวอย่างขนาด 500 1,000 และ 3,000

คน ในการศึกษาครั้งที่ 2 พบว่า เมื่อใช้กลุ่มตัวอย่างขนาดเล็กไม่เกิน 1,000 คน วิธี MH และวิธี SIBTEST ให้ผลในด้านความผิดพลาดชนิดที่ 1 ใกล้เคียงกันในระดับ .05 และวิธี MH มีความผิดพลาดชนิดที่ 1 สูงกว่าวิธี SIBTEST เมื่อกลุ่มตัวอย่างมีขนาดใหญ่มากกว่า 1,000 คน

จากการศึกษางานวิจัยที่ผ่านมา เป็นกรณีศึกษาวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบสองค่ามักพบว่าวิธี MH ซึ่งเป็นวิธีการของสถิติสหพารามेटริกใช้ได้ดีกับกลุ่มตัวอย่างขนาดเล็ก แต่จากการศึกษาครั้งนี้พบว่า วิธี GMH ซึ่งเป็นวิธีการของสถิติสหพารามेटริกของกลุ่มที่ไม่ใช่ IRT เมื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่ากับกลุ่มตัวอย่างขนาดเล็ก (250 คน) แม้จะมีความผิดพลาดในการตรวจสอบ ต่ำ (7% – 10 %) แต่ความถูกต้องในการตรวจสอบมีค่าต่ำมาก (ไม่ถึง 15 %) ส่วนการใช้วิธี GMH กับกลุ่มตัวอย่างขนาดใหญ่ (1,000 คน) แม้จะมีความถูกต้องสูง (50 % ขึ้นไป) แต่มีความผิดพลาดสูงมาก (32% – 73%) ด้วย ส่วนกลุ่มตัวอย่างขนาดกลาง (500 คน) การใช้วิธี GMH ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบมีความถูกต้องในการตรวจสอบ เพียง 23% ถึง 38% มีความผิดพลาดประมาณ 16 เปอร์เซนต์ ทั้งนี้ อาจพิจารณาได้ว่าวิธี GMH มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่า ต่ำ ซึ่งสอดคล้องกับการศึกษาของ มิลเลอร์และสเปรย์ (Miller & Spray, 1993) ที่ศึกษาเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่า ด้วยวิธี LDFA กับวิธี GMH โดยใช้ข้อมูลผลการสอบการปฏิบัติวิชาคณิตศาสตร์ พบว่า วิธี GMH ขาดอำนาจในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบไม่สม่ำเสมอ และการศึกษาของเวลช์และฮูเวอร์ (Miller & Spray, 1993 : 108 ; citing Welch & Hoover, 1993) ที่พบว่า แม้แต่ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบสม่ำเสมอ วิธี GMH ก็มีอำนาจในการตรวจสอบต่ำด้วย และ เสนอแนะว่า วิธี GMH อาจจะได้เป็นวิธีการทางเลือกที่ดีที่สุดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในการตอบข้อสอบที่เป็นแบบหลายค่า จากการศึกษาของ อุตทาโรและมิลล์ซาป (Uttaro & Millsap, 1994 : 24-25) ได้ให้ความเห็นว่า การศึกษาความถูกต้องของวิธีการแมนเทิล-แฮนส์เซล ยังต้องศึกษาให้กว้างขวางกว่านี้และควรเป็นการศึกษาจากสถานการณ์จริง เพื่อพิจารณาความถูกต้องในการตรวจสอบและความผิดพลาดชนิดที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบภายใต้เงื่อนไขต่าง ๆ ในระดับที่ยอมรับได้

1.2 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี Polytomous SIBTEST

ความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี Polytomous SIBTEST ระหว่างกลุ่มตัวอย่างขนาดต่างกัน คือ 1,000 คน, 500 คนและ 250 คน พบว่า ขนาดกลุ่มตัวอย่างส่งผลต่อความถูกต้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี Polytomous SIBTEST แตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 เกือบทุกเงื่อนไข ยกเว้นเมื่อกลุ่มตัวอย่างขนาด 500 คนและ 250 คน เมื่อแบบทดสอบมีขนาด 40 ข้อและ 30 ข้อ และกลุ่มตัวอย่างขนาด 1,000 คน และ 500 คน เมื่อแบบทดสอบมีขนาด 30 ข้อที่ส่งผลต่อความถูกต้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี Polytomous SIBTEST แตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 นั่นคือ เมื่อกลุ่มตัวอย่างมีขนาดใหญ่ขึ้น ความถูกต้องในการตรวจสอบของวิธี Polytomous SIBTEST มีค่าเพิ่มมากขึ้นด้วย ซึ่งสอดคล้องกับการศึกษาของจิตติมา วรณศรี (2539) ที่เปรียบเทียบประสิทธิภาพของวิธีการ MH และวิธี SIBTEST พบว่า เมื่อกลุ่มตัวอย่างมีขนาด 200 และ 600 คน ทั้งสองวิธีสามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ถูกต้องร้อยละ 50 แต่ถากลุ่มตัวอย่างขนาด 1,000 คน สามารถตรวจสอบได้ถูกต้องร้อยละ 100 และการศึกษาของสิริรัตน์ วิภาสศิลป์ (2545) ที่เปรียบเทียบวิธี SIBTEST กับวิธี DFIT ในการตรวจสอบการทำหน้าที่

ที่ต่างกันของข้อสอบที่ใช้ความสามารถหลายมิติ พบว่า กลุ่มตัวอย่างขนาด 500 คนและ 1,000 คน ส่งผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี SIBTEST สูงกว่ากลุ่มตัวอย่างขนาด 50 คน, 100 คนและ 200 คน

ความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี Polytomous SIBTEST ระหว่างกลุ่มตัวอย่างขนาดต่างกัน คือ 1,000 คน, 500 คนและ 250 คน พบว่า กลุ่มตัวอย่างขนาด 1,000 คน ส่งผลให้วิธี Polytomous SIBTEST มีความผิดพลาดในการตรวจสอบสูงกว่ากลุ่มตัวอย่างขนาด 500 คน และ 250 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 เกือบทุกเงื่อนไขความยาวของแบบทดสอบ ยกเว้นกรณีแบบทดสอบขนาด 20 ข้อ ที่กลุ่มตัวอย่างขนาด 1,000 คน ส่งผลให้วิธีการ Polytomous SIBTEST มีความผิดพลาดในการตรวจสอบสูงกว่ากลุ่มตัวอย่างขนาด 500 คน อย่างไม่มีนัยสำคัญ ส่วนกลุ่มตัวอย่างขนาด 500 คน ส่งผลให้วิธีการ Polytomous SIBTEST มีความผิดพลาดในการตรวจสอบสูงกว่ากลุ่มตัวอย่างขนาด 250 คน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 เมื่อแบบทดสอบมีขนาด 40 ข้อและ 30 ข้อ และอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เมื่อแบบทดสอบมีขนาด 20 ข้อ นั่นคือ เมื่อกลุ่มตัวอย่างมีขนาดใหญ่ขึ้นความผิดพลาดในการตรวจสอบของวิธี Polytomous SIBTEST มีค่าเพิ่มมากขึ้นด้วย

สรุปได้ว่า ขนาดของกลุ่มตัวอย่าง ส่งผลต่อประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี Polytomous SIBTEST ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่า นั่นคือ เมื่อกลุ่มตัวอย่างมีขนาดเพิ่มขึ้น มีผลทำให้ความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันมีค่าสูงขึ้น แต่ก็มีความผิดพลาดในการตรวจสอบมีค่าสูงขึ้นด้วย ซึ่งสอดคล้องกับการศึกษาที่ใช้ข้อสอบที่มีการให้คะแนนแบบสองค่า

2. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีต่างกัน

เมื่อแบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ กลุ่มตัวอย่างขนาด 1,000 500 และ 250 คน ผลตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GPCM พบข้อสอบที่แสดงการทำหน้าที่ต่างกันจำนวน 8 ข้อ วิธี GMH มีความถูกต้องในการตรวจสอบประมาณร้อยละ 13.75 ถึง 57.5 และมีแนวโน้มที่มีความถูกต้องมากขึ้นเมื่อกลุ่มตัวอย่างมีขนาดเพิ่มขึ้น ส่วนความผิดพลาดในการตรวจสอบมีค่าประมาณร้อยละ 7.82 ถึง 31.88 และมีแนวโน้มที่ความผิดพลาดสูงขึ้นเมื่อกลุ่มตัวอย่างมีขนาดเพิ่มขึ้น วิธี Polytomous SIBTEST มีความถูกต้องในการตรวจสอบประมาณร้อยละ 12.5 ถึง 52.5 และมีแนวโน้มที่ความถูกต้องมีมากขึ้นเมื่อกลุ่มตัวอย่างมีขนาดเพิ่มขึ้น ความผิดพลาดในการตรวจสอบมีค่าประมาณร้อยละ 21.88 ถึง 46.25 มีแนวโน้มที่ความผิดพลาดมีสูงขึ้นเมื่อกลุ่มตัวอย่างมีขนาดเพิ่มขึ้น

เมื่อแบบทดสอบประกอบด้วยข้อสอบ 30 ข้อ กลุ่มตัวอย่างขนาด 1,000 คน, 500 คนและ 250 คน เมื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GPCM พบข้อสอบที่แสดงการทำหน้าที่ต่างกันจำนวน 6 ข้อ วิธี GMH มีความถูกต้องในการตรวจสอบประมาณร้อยละ 13.34 ถึง 66.67 มีแนวโน้มที่ความถูกต้องมีมากขึ้นเมื่อกลุ่มตัวอย่างมีขนาดเพิ่มขึ้น ส่วนความผิดพลาดในการตรวจสอบมีค่าประมาณร้อยละ 10.00 ถึง 36.25 มีแนวโน้มที่ความผิดพลาดมีสูงขึ้นเมื่อกลุ่มตัวอย่างมีขนาดเพิ่มขึ้น วิธี Polytomous SIBTEST มีความถูกต้องในการตรวจสอบประมาณร้อยละ 31.63 ถึง 63.34 และมีแนวโน้มที่ความถูกต้องมีมากขึ้นเมื่อกลุ่มตัวอย่างมีขนาดเพิ่มขึ้น ความผิดพลาดในการตรวจสอบมีค่าประมาณร้อยละ 17.08 ถึง 56.67 มีแนวโน้มที่ความผิดพลาดสูงขึ้นเมื่อกลุ่มตัวอย่างมีขนาดเพิ่มขึ้น เช่นเดียวกัน

เมื่อแบบทดสอบประกอบด้วยข้อสอบ 20 ข้อ กลุ่มตัวอย่างขนาด 1,000 คน, 500 คนและ 250 คน เมื่อตรวจสอบสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GPCM พบข้อสอบที่แสดงการทำหน้าที่ต่างกัน จำนวน 12 ข้อ วิธี GMH มีความถูกต้องในการตรวจสอบประมาณร้อยละ 10.83 ถึง 51.67 และมีแนวโน้มที่ความถูกต้องมีมากขึ้นเมื่อกลุ่มตัวอย่างมีขนาดเพิ่มขึ้น ส่วนความผิดพลาดในการตรวจสอบมีค่าประมาณร้อยละ 7.5 ถึง 72.5 และมีแนวโน้มที่ความผิดพลาดมีสูงขึ้นเมื่อกลุ่มตัวอย่างมีขนาดเพิ่มขึ้น วิธี Polytomous SIBTEST มีความถูกต้องในการตรวจสอบประมาณร้อยละ 27.5 ถึง 69.17 และมีแนวโน้มที่ความถูกต้องมีมากขึ้นเมื่อกลุ่มตัวอย่างมีขนาดเพิ่มขึ้น ความผิดพลาดในการตรวจสอบมีค่าประมาณร้อยละ 18.75 ถึง 46.25 มีแนวโน้มที่ความผิดพลาดมีสูงขึ้นเมื่อกลุ่มตัวอย่างมีขนาดเพิ่มขึ้น เช่นเดียวกัน

จะเห็นได้ว่า ความถูกต้องในการตรวจสอบและความผิดพลาดในการตรวจสอบ ด้วยวิธี GMH และวิธี Polytomous SIBTEST มีความผันแปรค่อนข้างเป็นระบบในเงื่อนไขของขนาดของกลุ่มตัวอย่าง และความยาวของแบบทดสอบ นอกจากนั้น ความถูกต้องในการตรวจสอบและความผิดพลาดในการตรวจสอบของวิธี GMH และวิธี Polytomous SIBTEST มีค่าค่อนข้างใกล้เคียงกัน ในแต่ละขนาดของกลุ่มตัวอย่าง และขนาดของแบบทดสอบ ความถูกต้องในการตรวจสอบของวิธี GMH และวิธี Polytomous SIBTEST มีค่าสูงขึ้นตามขนาดของกลุ่มตัวอย่าง แต่ก็มีผลผิดพลาดในการตรวจสอบสูงขึ้นตามขนาดของกลุ่มตัวอย่างด้วย ทั้งนี้อาจมีปัจจัยหลายประการที่อาจเป็นสาเหตุ ดังนี้

1. ลักษณะของข้อสอบ เป็นปัจจัยที่มีผลต่อความถูกต้องและความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (Narayanan & Swaminathan. 1994) ข้อสอบวิชาคณิตศาสตร์ เรื่อง วงกลม ที่ผู้วิจัยสร้างขึ้นนั้น มีลักษณะเป็นข้อสอบปรนัยแบบเลือกตอบแบบถูกผิด (Multiple True-False Item) ชนิด 5 ตัวเลือก จำนวน 40 ข้อ ที่ผู้วิจัยสร้างให้แสดงการทำหน้าที่ต่างกันของข้อสอบที่เอื้อต่อเพศชาย จำนวน 8 ข้อ คือ ข้อ 1, 4, 17, 24, 28, 29, 39 และ 40 โดยกำหนดให้ผู้สอบต้องพิจารณาและตอบทุกตัวเลือก การศึกษาครั้งนี้พบประเด็นที่น่าพิจารณา ดังนี้

1.1. การกำหนดเงื่อนไขในการสร้างแบบทดสอบ ที่แสดงการทำหน้าที่ต่างกันของข้อสอบที่เอื้อต่อเพศชาย โดยกำหนดให้มีลักษณะของโจทย์ปัญหาที่ซับซ้อน โจทย์ปัญหาที่พิจารณาแล้วต้องแปลงให้เป็นภาพ ต้องใช้ความสามารถทางมิติสัมพันธ์ในการมองรูปทรง การคาดคะเนความยาว ระยะทาง การเชื่อมโยงสิ่งที่กำหนด และต้องใช้ความพยายามในการค้นหาคำตอบ ลักษณะของตัวเลือกที่ต้องวิเคราะห์ซับซ้อนในการค้นหาคำตอบ และจากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง พบว่า ลักษณะของข้อสอบคณิตศาสตร์ที่เพศชายจะทำได้ดีกว่าเพศหญิง ได้แก่ ข้อสอบเรขาคณิตและพีชคณิต ที่มีเหตุผลเชิงคณิตศาสตร์, งานที่มีลักษณะเป็นการแก้ปัญหา (Problem Solving) และงานที่ใช้ความคิดในขั้นสูง (higher level cognitive), ข้อสอบเกี่ยวกับอัตราส่วน (ratio), สัดส่วน (proportions) หรือ เปอร์เซนต์ (percent) และข้อสอบที่มีตาราง, กราฟ หรือ ภาพ ส่วนลักษณะของข้อสอบคณิตศาสตร์ที่เพศหญิงจะทำได้ดีกว่าเพศชาย ได้แก่ การคิดคำนวณ, พีชคณิต, ข้อสอบที่ใช้สัญลักษณ์และข้อสอบคณิตศาสตร์ที่เป็นนามธรรม (abstract mathematics) และงานที่มีลักษณะคล้ายคลึงกับในตำรา ในการสร้างข้อสอบจึงเป็นการสร้างตามแนวคิดและหลักการที่ได้ศึกษามา และการพิจารณาของผู้เชี่ยวชาญ ก็เป็นการพิจารณาไปตามเกณฑ์ที่ผู้วิจัยกำหนดเงื่อนไขต่าง ๆ ตามหลักการที่ศึกษาดังกล่าวข้างต้น แต่ในสภาพจริงของการเรียนการสอนในประเทศไทย อาจมีปัญหาและตัวแปรที่เป็นปัจจัยให้ผลการศึกษา ไม่เป็นไปตามหลักการที่อ้างอิงจากการศึกษาของต่างประเทศ เช่น พฤติกรรมการสอนของครู พฤติกรรมการเรียนของนักเรียน เป็นต้น และจากข่าวคณิตศาสตร์ของวารสาร การศึกษาวิทยาศาสตร์ คณิตศาสตร์และเทคโนโลยี (2545) ได้กล่าวถึง ความเชื่อยอดนิยม (ที่ผิด ๆ) เกี่ยว

กับการเรียนคณิตศาสตร์ ที่ว่า "โดยธรรมชาติผู้ชายมีความถนัดในวิชาคณิตศาสตร์มากกว่าผู้หญิง" แต่ข่าวนี้ก็ไม่มีข้อมูลยืนยันที่ผ่านการศึกษาย่างชัดเจน นั่นคือ สิ่งที่ต้องศึกษาและค้นหา ต่อไป

1.2 สัดส่วนของเวลากับจำนวนข้อสอบ คือ การกำหนดให้นักเรียนทำข้อสอบปรนัยแบบเลือกตอบแบบถูกผิด ชนิด 5 ตัวเลือก จำนวน 40 ข้อ ที่ต้องพิจารณาและตอบทุกตัวเลือก โดยกำหนดเวลาในการทำข้อสอบ 100 นาที นั้น นักเรียนจะต้องคิดคำนวณและตอบข้อสอบทั้งหมดจำนวน 200 ข้อย่อย ถ้าคิดเวลาในการทำข้อสอบโดยเฉลี่ยข้อละ 2 นาที อีกทั้งลักษณะของข้อสอบเป็นการวัดกระบวนการค้นหาข้อสรุปเป็นขั้นตอนโดยต้องใช้ความรู้ความจำเกี่ยวกับนิยาม ทฤษฎี เป็นพื้นฐาน ใช้ทักษะในการคิดคำนวณ การแก้ปัญหาและการเชื่อมโยงความสัมพันธ์ ถ้าพิจารณาเปรียบเทียบกับเวลาที่ใช้ในการสอบกับจำนวนข้อสอบที่ต้องทำข้อสอบทั้งหมดแล้ว อาจเป็นสาเหตุที่ทำให้นักเรียนเกิดความเหนื่อยล้า เบื่อหน่าย และทอดถอนในการทำข้อสอบได้ ซึ่งมีผลต่อคะแนนผลการตอบข้อสอบ

1.3 รูปแบบการตอบข้อสอบ ในสภาพการสอบวัดปกติ ครูผู้สอนมักใช้ข้อสอบปรนัยเลือกตอบธรรมดาเป็นส่วนใหญ่ ดังนั้น การใช้ข้อสอบปรนัยเลือกตอบแบบถูกผิดที่ต้องพิจารณาทุกตัวเลือก อาจทำให้นักเรียนไม่มีความถนัด ไม่คุ้นเคยกับการทำข้อสอบ ซึ่งอาจมีผลต่อคะแนนผลการตอบข้อสอบ

1.4 คุณภาพและลักษณะทางสถิติของข้อสอบวิชาคณิตศาสตร์ที่ผู้วิจัยสร้างขึ้นนั้นมีค่าความยากอยู่ในช่วง .50 ถึง .80 (ยากปานกลาง ถึงค่อนข้างง่าย) และมีค่าอำนาจจำแนกโดยเฉลี่ย .38 (.20 - .61 ถือว่าค่อนข้างต่ำ) และเมื่อพิจารณาค่าสถิติพื้นฐานของแบบทดสอบขนาด 40 ข้อ พบว่า มีค่าเฉลี่ย 116.71 จากคะแนนเต็ม 200 คะแนน ค่าส่วนเบี่ยงเบนมาตรฐาน 17.23 ค่าความเบ้ (Skewness) 1.6 นั่นคือ มีลักษณะเบ้ทางบวกเล็กน้อย (Positive skewness) แสดงว่า ผลการสอบส่วนใหญ่มีค่าคะแนนต่ำกว่าค่าเฉลี่ยเล็กน้อย และค่าความโด่ง (Kurtosis) 3.44 แสดงว่า มีลักษณะการแจกแจงความถี่ในลักษณะค่อนข้างโด่ง นั่นคือ การแจกแจงของคะแนนของกลุ่มตัวอย่างมีลักษณะเบ้ทางบวกแทนที่จะมีการแจกแจงปกติ จากการศึกษาของโมนาโค (Monaco, 1997) โดยใช้ข้อมูลจำลอง พบว่า เมื่อการแจกแจงพารามิเตอร์ความสามารถมีความเบ้ ทำให้ความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ลดลง และการศึกษาของ โรเจอร์ และ สวามินาธาน (Roger & Swaminathan, 1993) พบว่า เมื่อลักษณะของข้อสอบมีค่าอำนาจจำแนกสูงและค่าความยากปานกลางจะมีผลทำให้อำนาจในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบสมมาตร และแบบไม่สมมาตรของทั้งวิธี MH และวิธี LR มีค่าสูงสุด แต่เมื่อลักษณะของข้อสอบมีค่าอำนาจจำแนกต่ำ และมีค่าความยากปานกลาง จะมีผลทำให้อำนาจในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบไม่สมมาตรของวิธี MH มีค่าต่ำสุด และนารานายันและสวามินาธาน (Narayanan & Swaminathan, 1994) พบว่า ข้อสอบที่มีค่าอำนาจจำแนกสูงจะทำให้มีอำนาจในการตรวจสอบ สูง คือ ตรวจสอบพบข้อสอบที่ทำหน้าที่ต่างกันได้ถูกต้องมากที่สุด

2. การใช้คะแนนรวมเป็นเกณฑ์ในการจับคู่กลุ่มผู้สอบมีผลต่อประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี GMH และวิธี Polytomous SIBTEST พบว่า ทั้งสองวิธีมีความถูกต้องในการตรวจสอบปานกลางและมีค่าความผิดพลาดในการตรวจสอบสูง ทั้งนี้ อาจเนื่องจากผู้วิจัยใช้การวิเคราะห์แบบขั้นตอนเดียว คือ ใช้คะแนนรวมของแบบทดสอบที่มีข้อสอบที่ทำหน้าที่ต่างกันรวมอยู่ด้วยเป็นเกณฑ์ในการจับคู่ของกลุ่มผู้สอบ ทำให้เกณฑ์มีความบิดเบือนไปจากความเป็นจริง (Holland & Thayer, 1988) ซึ่งการศึกษาของทองอยู่ สาระ (2543) พบว่า การวิเคราะห์แบบสองขั้นตอน เพื่อทำให้คะแนนรวมที่ใช้เป็นเกณฑ์ในการจับคู่ผู้สอบมีความบริสุทธิ์แล้ว ความผิดพลาดในการ

ตรวจสอบระหว่างวิธี MH และวิธี LR เมื่อใช้กลุ่มตัวอย่างทุกขนาด มีค่าไม่แตกต่างกัน และมีความแม่นยำในการตรวจสอบสูง

3. ขนาดของกลุ่มตัวอย่าง เป็นตัวแปรหนึ่งที่มีผลกระทบต่อประสิทธิภาพในการตรวจสอบ จากการศึกษาวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบสองค่า ที่เป็นวิธีการที่ใช้สถิติ้นพาราเมตริก อาจใช้กลุ่มตัวอย่างขนาดเล็กได้ กลุ่มตัวอย่างขนาดเล็กที่สุดที่ควรใช้ คือ 100 คน (Stout. et al. 1997 : 199) การศึกษาของสวามินาธานและโรเจอร์ (Swaminathan & Rogers. 1990) พบว่า เมื่อใช้กลุ่มตัวอย่างขนาด 250 คน วิธี MH และวิธี LR มีความถูกต้องแม่นยำในการตรวจสอบการทำหน้าที่ต่างกันแบบสม่ำเสมอได้ 75% และเมื่อใช้กลุ่มตัวอย่าง 500 คน ทั้งสองวิธีมีความถูกต้องแม่นยำในการตรวจสอบการทำหน้าที่ต่างกันแบบสม่ำเสมอได้ถึง 100% สำหรับความผิดพลาดชนิดที่ 1 พบว่า วิธี MH มีความผิดพลาดร้อยละ 1 และจากการศึกษาของรูสโซและสตาต์ (Roussos & Stout. 1996 : 215-230) ที่ศึกษาผลของการใช้กลุ่มตัวอย่างขนาดเล็ก ที่มีต่อความผิดพลาดชนิดที่ 1 โดยใช้กลุ่มตัวอย่างขนาด 100 คน, 200 คน, 500 คนและ 1,000 คน พบว่า ยกเว้นกลุ่มตัวอย่างขนาด 100 คนแล้ว ความผิดพลาดชนิดที่ 1 ของทั้งสองวิธีแตกต่างกันอย่างไม่มีนัยสำคัญ และพบว่า วิธี SIBTEST มีความผิดพลาดชนิดที่ 1 สูงกว่าวิธี MH เล็กน้อย แต่เมื่อใช้กลุ่มตัวอย่างขนาดไม่เกิน 1,000 คน มีแนวโน้มว่าความผิดพลาดสูงขึ้น เมื่อกลุ่มตัวอย่างมีขนาดใหญ่ขึ้นทั้งสองวิธี และการศึกษาของจิตติมา วรณศรี (2539) ที่เปรียบเทียบประสิทธิภาพของวิธี MH และวิธี SIBTEST พบว่า เมื่อกลุ่มตัวอย่างขนาด 200 คนและ 500 คน ทั้งสองวิธีสามารถตรวจสอบการทำหน้าที่ต่างกันได้ 50% แต่ถ้ากลุ่มตัวอย่างมีขนาด 1,000 คน สามารถตรวจสอบการทำหน้าที่ต่างกันได้ถูกต้อง 100% และสิริรัตน์ ภิลาสศิลป์ (2545) พบว่า กลุ่มตัวอย่างขนาด 50 คน, 100 คน และ 200 คน ส่งผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี SIBTEST ไม่แตกต่างกัน แต่กลุ่มตัวอย่างขนาด 500 และ 1,000 คน ส่งผลต่อความถูกต้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี SIBTEST สูงกว่ากลุ่มตัวอย่างขนาด 50 คน, 100 คนและ 200 คน แต่ความผิดพลาดในการตรวจสอบก็สูงกว่าด้วย แต่การศึกษาของกาญจนา วัจนสุนทร ((2538) ที่ศึกษาเพื่อพัฒนาเกณฑ์ตัดสินข้อสอบลำเอียงทางเพศ พบว่า วิธี MH และวิธี SIBTEST ควรใช้กลุ่มตัวอย่าง 600 คนขึ้นไป จากการศึกษาครั้งนี้ พบว่า ทั้งสองวิธี ขนาดของกลุ่มตัวอย่าง มีผลต่อประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ แตกต่างกัน คือ เมื่อกลุ่มตัวอย่างมีขนาดเพิ่มขึ้น มีผลทำให้ความถูกต้องในการตรวจสอบสูงขึ้น แต่ความผิดพลาดในการตรวจสอบก็สูงขึ้นด้วย ผู้วิจัยเห็นว่า เมื่อแบบทดสอบมีขนาด 20 ข้อ, 30 ข้อและ 40 ข้อ และกลุ่มตัวอย่างขนาด 250 คน, 500 คนและ 1,000 คน เมื่อตรวจสอบด้วยวิธี GMH มีความถูกต้องในการตรวจสอบโดยเฉลี่ยร้อยละ 12.64 29.03 และ 58.61 ตามลำดับ ส่วนวิธี Polytomous SIBTEST มีความถูกต้องในการตรวจสอบโดยเฉลี่ยร้อยละ 23.88 41.25 และ 61.67 ตามลำดับ แต่เมื่อพิจารณาความผิดพลาดในการตรวจสอบ พบว่า กลุ่มตัวอย่างขนาด 250 คน, 500 คนและ 1,000 คน เมื่อตรวจสอบด้วยวิธี GMH มีความผิดพลาดในการตรวจสอบโดยเฉลี่ยร้อยละ 8.44 16.22 และ 46.88 ตามลำดับ ส่วนวิธี Polytomous SIBTEST มีความผิดพลาดในการตรวจสอบโดยเฉลี่ยร้อยละ 19.24 32.19 และ 49.72 ตามลำดับ นั่นคือ ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่าด้วยวิธี GMH และวิธี Polytomous SIBTEST ควรใช้กลุ่มตัวอย่างขนาด 1,000 คน จะทำให้มีความถูกต้องในการตรวจสอบโดยเฉลี่ยร้อยละ 58.61 และ 61.67 ตามลำดับ ซึ่งเป็นความถูกต้องที่พอเชื่อถือได้ แม้จะมีความผิดพลาดสูงด้วยก็ตาม

4. ความยาวของแบบทดสอบ เป็นตัวแปรหนึ่งที่มีผลกระทบต่อประสิทธิภาพในการตรวจสอบ การทำหน้าที่ต่างกันของข้อสอบ เนื่องจากการตรวจสอบตามแนวคิดของวิธี Mantel-Haenszel และแนวคิด ของวิธี SIBTEST ที่ใช้คะแนนรวมจากผลการตอบข้อสอบของผู้สอบแทนความสามารถของผู้สอบ แล้วย่อม ส่งผลต่อความน่าเชื่อถือมากกว่า นั่นคือ เมื่อแบบทดสอบที่มีจำนวนข้อเพิ่มมากขึ้นแล้ว ความเชื่อมั่นจะมีค่า เพิ่มขึ้น และมีผลทำให้อำนาจในการตรวจสอบมีค่าเพิ่มมากขึ้น (Narayanan & Swaminathan. 1996) และการศึกษาของสวามินาธานและโรเจอร์ส (Swaminathan & Rogers. 1990) ที่พบว่า เมื่อใช้แบบทดสอบ ที่มีความยาวมากขึ้น จะมีส่งผลทำให้ความถูกต้องในการตรวจสอบ ของวิธี MH และวิธี LR มีค่ามากขึ้น ยกเว้นในกรณีที่เป็นแบบไม่สม่ำเสมอของวิธี MH และการศึกษาของจิตมา วรณศรี (2539) ที่พบว่า เมื่อใช้ แบบทดสอบขนาด 60 ข้อ จะมีผลทำให้วิธี MH และวิธี SIBTEST มีประสิทธิภาพในการตรวจสอบดีที่สุด ส่วน ผลการศึกษาของอุตทาโรและมิลล์แซป (Uttaro & Millsap. 1994) ที่พบว่า เมื่อใช้แบบทดสอบที่มีความ ยาวมากขึ้นแล้ว ความผิดพลาดชนิดที่ 1 ของวิธี MH จะมีค่าลดลง ส่วนกลุ่มที่ศึกษาแล้วพบว่า ความยาว ของแบบทดสอบ ไม่มีผลต่อประสิทธิภาพของการตรวจสอบ เช่น การศึกษาของโรเจอร์สและสวามินาธาน (Rogers & Swaminathan. 1993) ที่พบว่า ความยาวของแบบทดสอบ ไม่มีผลต่อความถูกต้องในการตรวจ สอบของวิธี MH และวิธี LR ยกเว้นในกรณีที่เป็นแบบไม่สม่ำเสมอของวิธี LR และการศึกษาของกาญจนา วัจนสุนทร (2538) ที่พบว่า ความยาวของแบบทดสอบ ไม่มีผลกระทบต่อความถูกต้องในการตรวจสอบและ ความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี SIBTEST และวิธี MH และการศึกษา ของสิริรัตน์ วิภาสศิลป์ (2545) ที่พบว่า ความยาวของแบบทดสอบ ไม่ส่งผลต่อความถูกต้องในการตรวจสอบ และความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี SIBTEST และวิธี DFIT ซึ่งการ ศึกษาดังกล่าวข้างต้น เป็นการศึกษาในกรณีที่ข้อมูลเป็นผลการตอบข้อสอบ ที่มีการให้คะแนนแบบสองค่า แต่จากการศึกษาครั้งนี้ เป็นการศึกษาในกรณีที่ข้อมูลเป็นผลการตอบข้อสอบที่มีการให้คะแนนแบบหลายค่า พบว่า เมื่อกลุ่มตัวอย่างมีขนาด 500 คนและ 250 คน แบบทดสอบที่มีขนาด 40 ข้อ, 30 ข้อและ 20 ข้อ ส่งผลต่อความถูกต้องในการตรวจสอบและความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มี การให้คะแนนแบบหลายค่าของทั้งสองวิธี ไม่แตกต่างกัน แสดงว่า การเพิ่มจำนวนข้อสอบจาก 20 ข้อเป็น 30 ข้อหรือ 40 ข้อ และการเพิ่มจำนวนข้อสอบจาก 30 ข้อเป็น 40 ข้อ เมื่อกลุ่มตัวอย่างมีขนาด 500 คน และ 250 คน ไม่มีผลทำให้ความถูกต้องในการตรวจสอบและความผิดพลาดในการตรวจสอบการทำหน้าที่ ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่าของทั้งสองวิธี มีค่าสูงขึ้น ส่วนในกลุ่มตัวอย่างขนาด 1,000 คน แบบทดสอบขนาด 40 ข้อ, 30 ข้อและ 20 ข้อ ส่งผลต่อความถูกต้องในการตรวจสอบและความผิดพลาด ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่าของทั้งสองวิธี แตกต่างกัน อย่างไม่เป็นระบบ แต่อาจมีผลร่วมกับกลุ่มตัวอย่างขนาดใหญ่ เนื่องจากในกลุ่มตัวอย่างขนาด 1,000 คน เมื่อเพิ่มความยาวของแบบทดสอบจาก 20 ข้อเป็น 30 ข้อและ 40 ข้อ มีผลต่อความผิดพลาดในการตรวจ สอบของสองวิธี อย่างมีนัยสำคัญ หรือ เมื่อเพิ่มความยาวของแบบทดสอบจาก 20 ข้อเป็น 30 ข้อและ 40 ข้อ ส่งผลให้ความผิดพลาดในการตรวจสอบมีค่าสูงขึ้นอย่างมีนัยสำคัญ ทั้งนี้ อาจเป็นไปได้ว่า ความยาว ของแบบทดสอบไม่ส่งผลต่อความถูกต้องในการตรวจสอบและความผิดพลาดในการตรวจสอบการทำหน้าที่ ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่าของทั้งสองวิธี โดยตรง แต่อาจมีผลร่วมกันกับขนาดของ กลุ่มตัวอย่าง

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1.1 การศึกษาครั้งนี้พบว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์ เรื่อง วงกลม เมื่อแบบทดสอบประกอบด้วยข้อสอบ 40 ข้อ, 30 ข้อและ 20 ข้อ ด้วยวิธี GMH และวิธี Polytomous SIBTEST ให้ผลในการตรวจสอบที่แตกต่างกันในเงื่อนไขของขนาดของกลุ่มตัวอย่างที่ต่างกัน นั่นคือ ขนาดของกลุ่มตัวอย่างมีผลต่อประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ดังนั้นในการศึกษาวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่านั้น ควรใช้กลุ่มตัวอย่างขนาดกลางขึ้นไป (ไม่ควรต่ำกว่า 500 คน) หรือกลุ่มตัวอย่างขนาดใหญ่ จะทำให้ผลการตรวจสอบการทำหน้าที่ต่างกันมีความถูกต้องในการตรวจสอบสูง แม้จะมีความผิดพลาดในการตรวจสอบสูงด้วยก็ตาม

1.2 การศึกษาครั้งนี้ พบว่า การสร้างแบบทดสอบที่ข้อสอบมีลักษณะเป็นข้อสอบปรนัยแบบเลือกตอบแบบถูกผิด (Multiple True-False Item) ชนิด 5 ตัวเลือก นั้น ไม่ควรมีจำนวนข้อมากนัก หรือ อาจลดจำนวนตัวเลือกลงบ้าง เพราะเมื่อกำหนดให้นักเรียนต้องพิจารณาและตอบทุกตัวเลือก นั่นคือ นักเรียนต้องทำข้อสอบเป็นจำนวนเท่ากับ จำนวนข้อสอบ คูณกับ จำนวนตัวเลือก จากการศึกษาครั้งนี้ จะเห็นได้ว่าแบบทดสอบขนาด 20 ข้อ มีความถูกต้องในการตรวจสอบสูงกว่าแบบทดสอบขนาด 30 ข้อและ 40 ข้อ ในเกือบทุกขนาดกลุ่มตัวอย่าง โดยเฉพาะวิธี Polytomous SIBTEST ที่ศึกษากับกลุ่มตัวอย่างขนาด 1,000 คน จะมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่า ได้ดีที่สุด และมีประสิทธิภาพในการตรวจสอบสูงกว่าวิธี GMH อย่างชัดเจน

1.3 ควรใช้การวิเคราะห์แบบสองขั้นตอน เพื่อให้เกณฑ์การจับคู่กลุ่มผู้สอบมีความบริสุทธิ์ (purification of matching criterion) ในการศึกษาวิธี GMH และวิธี Polytomous SIBTEST เพื่อการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ จะทำให้ผลการตรวจสอบมีประสิทธิภาพมากขึ้น

1.4 การศึกษาในครั้งนี้ได้ข้อค้นพบว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่าด้วยวิธี GMH มีขั้นตอนในการวิเคราะห์ที่ยุ้งยาก ซับซ้อน และควรใช้กลุ่มตัวอย่างขนาดกลางหรือขนาดใหญ่ขึ้นไป ซึ่งไม่ค่อยสะดวกในทางปฏิบัติเหมือนวิธี MH เดิม และยังไม่มียุ้โปรแกรมสำเร็จรูปที่ช่วยในการวิเคราะห์ ส่วนวิธี Polytomous SIBTEST ได้ผลการตรวจสอบที่ใกล้เคียงกันและมีประสิทธิภาพในการตรวจสอบของสูงกว่าวิธี GMH เล็กน้อยนั้น การวิเคราะห์มีความสะดวกมาก ไม่ยุ่งยาก ซับซ้อน เนื่องจากมีโปรแกรมสำเร็จรูป DIFPACK ช่วยในการวิเคราะห์ แต่ต้องเสียค่าใช้จ่ายในการสั่งซื้อโปรแกรมจากต่างประเทศในราคาที่สูงมาก ดังนั้น ทางเลือกที่จะใช้วิธีการใดจึงขึ้นอยู่กับความพร้อมและความเหมาะสมต่าง ๆ

1.5 การศึกษาในครั้งนี้ได้ข้อค้นพบว่า ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่าด้วยวิธี GPCM ซึ่งเป็นวิธีการตามแนวคิดของทฤษฎี IRT นั้น แม้จะมีข้อจำกัดในเรื่องต้องใช้กลุ่มตัวอย่างขนาดใหญ่ มีความยุ่งยากหลายขั้นตอนในการวิเคราะห์ เช่นเดียวกับวิธี GMH แต่เมื่อพิจารณาถึงจุดเด่นของวิธีการนี้ ในด้านความไม่แปรเปลี่ยนของค่าพารามิเตอร์ของข้อสอบไปตามกลุ่มผู้สอบ และด้านความไวและความละเอียดในการตรวจสอบ อีกทั้งสามารถให้รายละเอียดของผลการตรวจสอบได้มาก ผู้วิจัยเห็นว่าเป็นวิธีการหนึ่งที่น่าจะนำมาใช้ได้ดี โดยเฉพาะกับหน่วยงานที่เกี่ยวข้องกับการสอบวัดทางการศึกษา

2. ข้อเสนอแนะในการวิจัยครั้งต่อไป

2.1 ควรศึกษาวิธีการตรวจสอบเดิม โดยใช้แบบทดสอบในเนื้อหาวิชาอื่นบ้าง หรือ ใช้เครื่องมือที่มีรูปแบบการตอบในลักษณะอื่นๆ เช่น การทดสอบการปฏิบัติ การเขียนเรียงความ ข้อสอบเขียนตอบ เป็นต้น ซึ่งข้อสอบประเภทนี้มีการให้คะแนนแบบหลายค่า และที่สำคัญ คือ มีคุณค่าของสารสนเทศของผลการตอบข้อสอบของผู้สอบที่น่าสนใจเป็นอย่างมาก

2.2 ควรศึกษาเพื่อเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ของหมวดข้อสอบ (Differential Bundle Functioning : DBF) และของแบบทดสอบ (Differential Test Functioning : DTF) ด้วยวิธี GMH และวิธี Polytomous SIBTEST และควรศึกษาเปรียบเทียบกับข้อมูลผลการตอบที่ใช้ความสามารถหลายมิติ (Multidimensional response data) ด้วย

บรรณานุกรม

บรรณานุกรม

ภาษาไทย

- กัลยา วานิชย์บัญชา. (2544). การใช้ SPSS for Windows ในการวิเคราะห์ข้อมูล. กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย.
- . (2544). การวิเคราะห์สถิติ : สถิติเพื่อการตัดสินใจ. กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย.
- กาญจนา วัฒนสุนทร. (2538). การพัฒนาเกณฑ์ตัดสินข้อสอบลำเอียงทางเพศ. วิทยานิพนธ์ ค.ด. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- เกริกชัย ฮาบเจริญ. (2525). การเปรียบเทียบวิธีการตอบและตรวจให้คะแนนแบบทดสอบปรนัยชนิดเลือกตอบที่มีลักษณะแตกต่างกัน. ปรินูญานิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- เกสร หว่างจิต. (2539). การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบสำหรับแบบสอบคัดเลือกระดับบัณฑิตศึกษา วิชาภาษาไทยและภาษาอังกฤษ ด้วยวิธีแมนเทิล-แฮนส์เซล. วิทยานิพนธ์ ค.ม. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- โกสุม กรทอง. (2545). “ข่าวคณิตศาสตร์.” ใน วารสารการศึกษาวิทยาศาสตร์ คณิตศาสตร์และเทคโนโลยี. 30(117) : 70.
- จิตติมา วรณศรี. (2539). การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทิล-แฮนส์เซลกับวิธีชิปเทสท์ เมื่อความยาวของแบบทดสอบ ขนาดกลุ่มตัวอย่างและอัตราส่วนของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบต่างกัน. วิทยานิพนธ์ ค.ม. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- จิตสุดา ธราพร. (2539). การเปรียบเทียบผลการวิเคราะห์ความลำเอียงของข้อสอบจากแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์และแบบทดสอบความเข้าใจทางภาษา. ปรินูญานิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- จักรกฤษณ์ สำราญใจ. (2538). ประสิทธิภาพเชิงสัมพัทธ์ของข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือกเทียบกับข้อสอบเลือกตอบแบบฉบับในแบบทดสอบวัดผลสัมฤทธิ์. วิทยานิพนธ์ ค.ด. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- เจน จันท์แสง. (2537). การวิเคราะห์ความลำเอียงของข้อสอบของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3. ปรินูญานิพนธ์ กศ.ม. (การวัดผลการศึกษา). พิษณุโลก : บัณฑิตวิทยาลัย มหาวิทยาลัยนเรศวร. ถ่ายเอกสาร.
- ชัชชัย เผ่าพงศ์. (2527). การวิเคราะห์ความลำเอียงของข้อสอบของแบบทดสอบวัดความถนัดทางการเรียนด้านคณิตศาสตร์และภาษาระดับชั้นมัธยมศึกษาตอนต้น. วิทยานิพนธ์ ศศ.ม. (การวัดผลการศึกษา). ขอนแก่น : มหาวิทยาลัยขอนแก่น. ถ่ายเอกสาร.
- ชลินทร์ ศิลปสมศักดิ์. (2544). การเปรียบเทียบความเชื่อมั่นของแบบทดสอบเลือกตอบถูกผิดที่มีการตรวจให้คะแนนตามกฎชิลลิ่ง โดยมีการจัดเรียงข้อสอบและกำหนดจุดชิลลิ่งต่างกัน. ปรินูญานิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- ชวลิต โพธิ์นคร. (2540). การศึกษาอิทธิพลของการให้คะแนนและการคัดเลือกข้อสอบที่มีต่อคุณภาพแบบทดสอบเลือกตอบ. ปรินูญานิพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

- ชวาล แพรัตกุล. (2530). เทคนิคการเขียนคำถามเลือกตอบ. กรุงเทพฯ : กิ่งจันทร์การพิมพ์.
- ญาณภัทร สีหะมงคล. (2540). การเปรียบเทียบความสอดคล้องของผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันระหว่างวิธี Lord's χ^2 วิธี Raju's Area Measures และวิธี Closed Interval Area. ปรินญานิพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- ทัศน์ย์ พิรมนตรี. (2530). การวิเคราะห์ความลำเอียงของแบบสอบวิชาคณิตศาสตร์โครงการตรวจสอบคุณภาพการศึกษาชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526. วิทยานิพนธ์ ค.ม. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- ทองอยู่ สาระ. (2543). การเปรียบเทียบอำนาจการตรวจสอบและการจำแนกผิดพลาด ในการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบสม่ำเสมอและแบบไม่สม่ำเสมอระหว่างวิธีแมนเทิล-แฮนส์เซลและวิธีการถดถอยโลจิสติกโดยใช้ความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่างต่างกัน. ปรินญานิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- นพมาศ พิพัฒน์สุข. (2541). การเปรียบเทียบประสิทธิภาพระหว่างวิธีแมนเทิล-แฮนส์เซล กับ วิธีถดถอยโลจิสติกในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อใช้เกณฑ์จับคู่เปรียบเทียบแตกต่างกันในแบบสอบชนิดพหุมิติ. วิทยานิพนธ์ ค.ม. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- นิรมล ชัยชวลิต. (2537). การเปรียบเทียบผลการวิเคราะห์ความลำเอียงของแบบทดสอบความเข้าใจในการอ่านภาษาไทย ตามทฤษฎีคลาสสิกอลที่ใช้วิธีวิเคราะห์ต่างกัน. ปรินญานิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- บุญเชิด ภิญโญนนท์พงษ์. (2544). การประเมินการเรียนรู้ที่เน้นผู้เรียนเป็นสำคัญ : แนวคิดและวิธีการ. กรุงเทพฯ : สำนักงานคณะกรรมการการศึกษาแห่งชาติ.
- พรรณี จิตมาศ. (2540). การเปรียบเทียบผลการวิเคราะห์ความลำเอียงข้อสอบ โดยใช้ขนาดกลุ่มผู้สอบและวิธีวิเคราะห์ต่างกัน. ปรินญานิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- พัชรี ปิยะภักดิ์. (2531). การวิเคราะห์ความลำเอียงของข้อสอบจากแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน. ปรินญานิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- รัชนีกร มุกดา. (2540). การเปรียบเทียบประสิทธิภาพระหว่างวิธีแมนเทิล-แฮนส์เซลกับวิธีถดถอยโลจิสติกในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอนุกรม ในการจัดการกลุ่มความสามารถ ค่าความยากของข้อสอบ และค่าอำนาจจำแนกของข้อสอบต่างกัน. วิทยานิพนธ์ ค.ม. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- เรวดี อินทะสระ. (2539). ผลการตรวจสอบความลำเอียงของข้อสอบต่อการศึกษาคำความเที่ยงตรงเชิงพยากรณ์ของแบบทดสอบคัดเลือกที่คิดคะแนนต่างกัน. ปรินญานิพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

- ล้วน สายยศ และอังคณา สายยศ. (2539). *เทคนิคการวัดผลการเรียนรู้*. กรุงเทพฯ : ชมรมเด็ก.
- วลีมาศ แซ่อึ้ง. (2543). *การเปรียบเทียบอำนาจการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอนุกรมระหว่างวิธีชิปเทสต์ปรับปรุงใหม่ วิธีชิปเทสต์วิธี MH และวิธี LR*. วิทยานิพนธ์ ค.ด. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- ศิริชัย กาญจนวาสี. (2543). *การประเมินการเรียนรู้: ข้อเสนอแนะเชิงนโยบาย*. กรุงเทพฯ : สำนักงานคณะกรรมการการศึกษาแห่งชาติ.
- ศิริเดช สุชีวะ. (2539, ก.ค.-ธ.ค.). "โมเดลการตอบสนองข้อสอบแบบพหุภาคและการประยุกต์ใช้กับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์." ใน *วิธีวิทยาการวิจัย*. 8(2) :17-30.
- สิริรัตน์ วิชาศิลป์. (2545). *การเปรียบเทียบวิธีชิปเทสต์และดีเอฟไอทีในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบ จากข้อมูลการตอบข้อสอบที่ใช้ความสามารถหลายมิติ*. ปรียญานิพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- สุพัฒน์ สุกมลสันต์. (2534). *การวิเคราะห์ความลำเอียงของข้อสอบภาษาอังกฤษเข้ามหาวิทยาลัย ปี 2531-2533*. สถาบันภาษา. กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- สุศักดิ์ อมรรตณศักดิ์. (2531). *การศึกษาเปรียบเทียบผลของวิธีวิเคราะห์ความลำเอียงของข้อสอบที่แตกต่าง 4 วิธี*. วิทยานิพนธ์ ค.ด. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- เสรี ชัดแฉ่ม. (2539). *การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันแบบไม่สม่ำเสมอของข้อสอบระหว่างวิธีแมนเทิล-แฮนส์เซลแบบปกติ กับวิธีแมนเทิล-แฮนส์เซลแบบแบ่งกลุ่มความสามารถผู้สอบและความยาวของข้อสอบ*. วิทยานิพนธ์ ค.ด. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- อารี วัชรโสติกุล. (2543). *การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยใช้รูปแบบและวิธีการต่างกัน*. ปรียญานิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

ภาษาอังกฤษ

- Agresti, A. (1990). *Categorical data analysis*. New York : Wiley.
- Albanese, M. A. & Sabers, D. L. (1988). "Multiple True-False Items: A Study of Interitem Correlation, Scoring Alternatives, and Reliability Estimation," *Journal of Educational Measurement*. 25(2) : 111-123.
- Allen, M. J. & Yen, W. M. (1979). *Introduction to Measurement Theory*. California : Brooks/Cole Publishing Company.
- Anderson, J. A. (1972). "Separate Sample Logistic Discriminant," *Biometrika*. 59(1): 19-35.
- Baker, F. B., Al-Karni, A. & Al-Dosary, I. M. (1991). "EQUATE: A Computer Program for the Test Characteristic Curve Method of IRT Equating," *Applied Psychological Measurement*. 15(1) : 78.

- Baker, F. B. (1993). "EQUATE2.0 : A Computer Program for the Characteristic Curve Method of IRT Equating," *Applied Psychological Measurement*. 17(1) : 20.
- Berk, R. A. (1982). *Handbook of Methods for Detecting Test Bias*. Baltimon and London Md : The Johns Hopkins University Press.
- Budgell, G. R., Raju, N. S., & Quartetti, D. A. (1995). "Analysis of Differential Item Functioning in Translated Assessment Instruments," *Applied Psychological Measurement*. 19(4): 309–321.
- Burrill, L. E. (1982). "Comparative Studies of Item Bias Method," In *Handbook of Methods for Detecting Test Bias*. Baltimon and London Md : The Johns Hopkins University Press.
- Camilli, G. & Shepard, L. A. (1994). *Methods for Identifying Biased Test Items*. California : Sage Publications, Inc.
- Chang, H. & Mazzeo, J. (1994). "The Unique Correspondence of the Item Response Function and Item Category Response Functions in Polytomously Scored Item Response Models," *Psychometrika*. 59(3): 391-404.
- Childs, R. A., & Chen, Wen-Hung. (1999). "Obtaining Comparable Item Parameter Estimates in MULTILOG and PARSCALE for Two Polytomous IRT Models," *Applied Psychological Measurement*. 23(4): 371–379.
- Clauser, B. E., Mazor, K. & Hambleton, R. K. (1991). "Influence of the Criterion Variable on the Identification of Differentially Functioning Test Items Using the Mantel-Haenszel Statistic," *Applied Psychological Measurement*. 15(4): 353–359.
- Cohen, A.S., Kim, S. H. & Subkoviak, M. J. (1991). "Influence of Prior Distributions on Detection of DIF," *Journal of Educational Measurement*. 28(1) : 49-59.
- Cox, T. F. & Ferry, G. (1991). "Robust Logistic Discriminant," *Biometrika*. 78(4): 841 – 849.
- Donoghue, J. R. (1994). "An Empirical Examination of the IRT Information of Polytomously Scored Reading Items Under the Generalized Partial Credit Model," *Journal of Educational Measurement*. 31(4) : 295-311.
- Dorans, N. J. & Kulick, E. (1986). "Demonstrating the Utility of the Standardization Approach to Assessing Unexpected Differential Item Performance on the Scholastic Aptitude Test," *Journal of Educational Measurement*. 23(4) : 355-368.
- Downing, S. M., et al. (1995). "Item Type and Cognitive Ability Measured: The Validity Evidence for Multiple True-False Items in Medical Specialty Certification," *Applied Measurement in Education*. 8(2) : 187-197.
- French, A. W. & Miller T. R. (1996). "Logistic Regresstion and Its Use in Detecting Differential Item Functioning in Polytomous Items," *Journal of Education Measurement*. 33(3): 315–332.
- Garner, M. & Engelhard, G. Jr. (1999). "Gender Differences in Performance on Multiple-Choice and Constructed Response Mathematics Items," *Applied Measurement in Education*. 12(1) : 29-51.

- Gipps, C. & Murphy, P. (1994). *A Fair Test? Assessment, Achievement and equity*. Buckingham : Open University Press.
- Green, B. F. (1994). "Differential Item Functioning : Techniques, findings and prospects," *Modern Theories of Measurement : Problem and Issue*. Ottawa : Edumetrics Research Group, University of Ottawa.
- Guilford, J. P. & Fruchter, B. (1978). *Fundamental Statistics in Psychology and Education*. McGraw-Hill: Singapore.
- Haladyna, T. M. (1994). *Developing and Validating Multiple-Choice Test Items*. Lawrence Erlbaum Associates, Hillsdale : New Jersey.
- Hambleton, R. K. & Rogers, H. J. (1989). "Detecting Potential Biased Test Items: Comparison of IRT Area and Mantel-Haenszel Methods," *Applied Measurement in Education*. 2(4):313-334.
- Hambleton, R.K. & Swaminathan, H. (1985). *Item Response Theory Principles and Applications*. Boston, Kluwer-Nijhoff Publishing.
- Hambleton, R.K., Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, California : Sage Publications, Inc.
- Hauck, W.W. (1983). A Note on Confidence Bands for the Logistic Response Curve. *The American Statistician*. 37, 158-160.
- Holland, P. W. & Thayer, D. T. (1988). "Differential Item Performance and the Mantel-Haenszel Procedure," In Wainer, Howard and Braun, Henry I. (Eds.) *Test Validity*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 129-146.
- Holland, P. W. & Wainer H. (1993). *Differential Item Functioning*. Hillsdale, New Jersey : Lawrence Erlbaum Associates.
- Hoskens, M. & Boeck, Paul De. (1995). "Componential IRT Models for Polytomous Items," *Journal of Educational Measurement*. 32(4) : 364 – 384.
- Huberty, C. J. (1994). *Applied Discriminant Analysis*. New York : A Wiley-Interscience Publication.
- Ip, E. H. (2001). "Testing for Local Dependency in Dichotomous and Polytomous Item Response Models," *Psychometrika*. 66(1): 109-132.
- Kim, S.H. & Cohen, A. S. (Spring, 1992). "Effect of Linking Methods on Detection of DIF," *Journal of Educational Measurement*. 29(1) : 51 – 66.
- Kim, S.H. & Cohen, A. S. (1995). "A Comparison of Lord's Chi-Square, Raju's Area Measures, and the Likelihood Ratio Test on Detection of Differential Item Functioning," *Applied Psychological Measurement*. 8(4) : 291-312.
- Kim, S.H., Cohen, A. S. & Kim, H. O. (1994). "An Investigation of Lord's Procedure for the Detection of Differential Item Functioning," *Applied Psychological Measurement*. 18(3) : 217-228.
- Li, H. & Stout, W. (1996). "A New Procedure for Detection of Crossing DIF," *Psychometrika*. 61(4) : 647-677.

- Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problem*. Hillsdale, New Jersey : Erlbaum.
- Lord, F. M. & Novick, M. R. (1968). *Statistical Theories of Mental Test Score*. Massachusetts : Addison Wesley Publishing Company.
- Mazor, K. M., Clauser, B. E. & Hambleton, R. K. (1992). "The Effect of Sample Size on the Functioning of the Mantel-Haenszel Statistic," *Educational and Psychological Measurement*. 52(3) : 443-451.
- Mazor, K. M. & Clauser, B. E. (1995). "Using Logistic Regression and Mantel-Haenszel with Multiple Ability Estimates to Detect Differential Item Functioning," *Journal of Educational Measurement*. 32(2) : 131-144.
- Mellenbergh, G. J. (1995). "Conceptual Notes on Models for Discrete Polytomous Item Responses," *Applied Psychological Measurement*. 19(1): 91-100.
- Miller, T. R. & Spray, J. A. (1993). "Logistic Discriminant Function Analysis for DIF Identification of Polytomously Scored Items," *Journal of Education Measurement*. 30(2) : 107 – 122.
- Millsap, R. E. & Everson, H. T. (1993). "Methodology Review:Statistical Approaches for Assessing Measurement Bias," *Applied Psychological Measurement*. 17(4) : 297 – 334.
- Monaco, Malina. (1997). "A Monte Carlo Study of Skewed Theta Distribution on DIF Indices," (online). Available <http://eric.syr.edu/plweb.../>
- Muraki, E. (1992). "A Generalized Partial Credit Model: Application of an EM Algorithm," *Applied Psychological Measurement*. 16(2): 159-176.
- Muraki, E. (1993). "Information Functions of the Generalized Partial Credit Model," *Applied Psychological Measurement*. 17(4): 351-363.
- Nandakumar, R. (1993). "Simultaneous DIF Amplification and Cancellation : Shealy-Stout's Test for DIF," *Journal of Educational Measurement*. 30(4) : 293-311.
- Narayanan, P. & Swaminathan H. (1994). "Performance of the Mantel-Haenszel and Simultaneous Item Bias Procedures for Detecting Differential Item Functioning," *Applied Psychological Measurement*. 18(4) : 315 – 328.
- Osterlind, S. J. (1985). *Test Item Bias*. California : Sage Publications, Inc.
- Potenza, M. T. & Dorans, N. J. (1995). "DIF Assessment for Polytomously Scored Items: A Framework for Classification and Evaluation," *Applied Psychological Measurement*. 19(1) : 23 – 37.
- Press, S. J. & Wilson, S. (1978). "Choosing Between Logistic Regression and Discriminant Analysis," *Journal of the American Statistical Association*. 73(364) : 699 – 705.
- Raju, S., et al. (1993). "An Empirical Comparison of the Area Methods : Lord's Chi-Square Test, and the Mantel-Haenszel Technique for Assessing Differential Item Function," *Educational and Psychological Measurement*. 53 : 301 – 314.

- Raju, S., van der Linden, J. & Fleer F. (1995). "IRT-based Internal Measures of Differential Functioning of Items and Tests," *Applied Psychological Measurement*. 15(4) : 361 – 373.
- Rogers, H. J. & Swaminathan H. (1993). "A Comparison of Logistic Regression and Mantel-Haenszel Procedures for Detecting Differential Item Functioning," *Applied Psychological Measurement*. 17(2) : 105 – 116.
- Roussos, L. A. & Stout, W. F. (1996). "Simulation Studies of the Effects of Small Sample Size and Studies Item Parameters on SIBTEST and Mantel-Haenszel Type I Error Performance," *Journal of Educational Measurement*. 33(2): 215-230.
- Sapra, S. K. (1991). "A Connection Between the Logit Model, Normal Discriminant Analysis, and Multivariate Normal Mixtures." *The American Statistician*. 45(4) : 265–268.
- Scheuneman, J. D. (1979). "A Method of Assessing Bias in Test Items," *Journal of Educational Measurement*. 16(3) : 143-152.
- Shealy, R. & Stout, W. F. (1993). "A Model-based Standardization Approach that Separates True Bias/DIF from Group Ability Differences and Detects Test Bias/DIF as well as Item Bias/DIF," *Psychometrika*. 58(2) : 159 – 194.
- Shealy, R. & Stout, W. F. (1991). *SIBTEST user Manual [Computer program manual]*. Champaign : University of Illinois, Department of Statistics.
- Seigel, S. (1956). *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill Kogakusha Tokyo.
- Stout, William. (2001). "Software : DIFPACK – Dimensionality-Based DIF Analysis Package," (online). Available <http://www.assess.com/Software/DIFPACKover.htm>
- Subkoviak, M. J., et al. (1984). "Empirical Comparison of Selected Item Bias Detection Procedures with Bias Manipulation," *Journal of Educational Measurement*. 21(1) : 49-58.
- Swaminathan, H. & Rogers, H. J. (1990). "Detecting Differential Item Functioning Using Logistic Regression Procedures," *Journal of Educational Measurement*. 27(4) : 361 – 370.
- Tate, R. (2000). "Robustness of the School-Level Polytomous IRT Model," *Educational and Psychological Measurement*. 60(1): 20-37.
- Thissen, D., Steinberg, L. & Wainer, H. (1993). "Detection of Differential Item Functioning Using the Parameters of Item Response Models," In Holland, P.W. and Wainer, H. (Eds.). *Differential Item Functioning*. Hillsdale, New Jersey : Lawrence Erlbaum Associates, Inc., 67-113.
- Uttaro, T. & Millsap, R. E. (1994). "Factors Influencing the Mantel-Haenszel Procedure in the Detection of Differential Item Functioning," *Applied Psychological Measurement*. 18(1):15-25.
- Welch, C., & Hoover, H.D. (1993). "Procedures for Extending Item Bias Techniques to Polytomously Scored Items," *Applied Measurement in Education*. 6, 1-9.
- Wilson, Jame.W. (1971). "An Evaluation of Learning in Secondary School Mathematics," in *Handbook on Formative Evaluation of Student Learning*. USA : McGraw-Hill.

- Wright, B. D. & Masters, G. N. (1982). *Rating Scale Analysis*. Mesa Press, Chicago.
- Wright, B. D. & Stone, M. H. (1979). *Best Test Design*. Mesa Press, Chicago.
- Yamane, Taro. (1974). *Statistics : an Introductory Analysis*. 2nd ed. New York : Harper.
- Zwick R. & Thayer, D. T. (1996). "Evaluating the Magnitude of Differential Item Functioning in Polytomous Items," *Journal of Educational and Behavioral Statistics*. 21(3): 187-201.
- Zwick, R., Thayer, D. T. & Mazzeo, J. (1997). "Descriptive and Inferential Procedures for Assessing Differential Item Functioning in Polytomous Items," *Applied Measurement in Education*. 10(4): 321-344.

ภาคผนวก

ภาคผนวก ก

รายชื่อผู้เชี่ยวชาญที่ตรวจสอบคุณภาพของแบบทดสอบ

1. อาจารย์พรณี อุ่นศรีเพ็ง
อาจารย์ 3 ระดับ 8 โรงเรียนบุญวาทย์วิทยาลัย จังหวัดลำปาง
2. อาจารย์กรรณิกา สุภาภา
อาจารย์ 2 ระดับ 7 โรงเรียนสตรีประเสริฐศิลป์ จังหวัดตราด
3. อาจารย์สุพร บุณยพงศ์
อาจารย์ 2 ระดับ 7 โรงเรียนเจ้าพระยาวิทยาคม กรุงเทพมหานคร
4. อาจารย์วรพร สกุลจิตรานนท์
อาจารย์ 2 ระดับ 7 โรงเรียนบ้านหมี่วิทยา จังหวัดลพบุรี
5. อาจารย์เบญจา เขียวสม
อาจารย์ 2 ระดับ 7 โรงเรียนปากน้ำหลังสวนวิทยา จังหวัดชุมพร

ภาคผนวก ข

รายชื่อโรงเรียนกลุ่มตัวอย่าง

ทดลองใช้เครื่องมือครั้งที่ 1

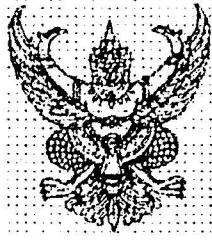
โรงเรียนอยุธยาวิทยาลัย
โรงเรียนจอมสุรางค์อุปถัมภ์

ทดลองใช้เครื่องมือครั้งที่ 2

โรงเรียนภาชี "สุนทรวิทยานุกูล"
โรงเรียนผักไห่ "สุทธาประมุข"
โรงเรียนบางซ้ายวิทยา
โรงเรียนบางปะอิน

ทดลองใช้เครื่องมือครั้งที่ 3

โรงเรียนท่าเรือ"นิตยานุกูล"
โรงเรียนบางปะอิน"ราชานุเคราะห์1"
โรงเรียนเสนา"เสนาประสิทธิ์"
โรงเรียนนครหลวง"อุดมรัชต์วิทยา"
โรงเรียนท่าช้างวิทยาคม
โรงเรียนบางไทรวิทยา
โรงเรียนบางปะหัน
โรงเรียนบ้านแพรกประชาสรรค์
โรงเรียนวัดโพธิ์ผักไห่(เวชพันธุ์อนุสรณ์)
โรงเรียนวิเชียรกลิ่นสุคนธ์อุปถัมภ์
โรงเรียนวังน้อย(พนมยงค์วิทยา)
โรงเรียนมหาราช"ประชานิยม"
โรงเรียนลาดชะโดสามัคคี
โรงเรียนปากกรานพิทยา



ที่ ศธ 0848 / 2662

สำนักงานสามัญศึกษาจังหวัดพระนครศรีอยุธยา
ศาลากลางจังหวัดพระนครศรีอยุธยา
ถนนเอเชีย อย 13000

16 กรกฎาคม 2545

เรื่อง ขอความร่วมมือเพื่อการวิจัย

เรียน ผู้บริหารโรงเรียนมัธยมศึกษา สังกัดกรมสามัญศึกษา จังหวัดพระนครศรีอยุธยา

สิ่งที่ส่งมาด้วย แบบสอบถามเพื่อการวิจัย

จำนวน ชุด

ด้วยนางสาวอุทัยวรรณ สายพัฒนะ นิสิตระดับปริญญาเอก สาขาการทดสอบและวัดผลการศึกษา มหาวิทยาลัยศรีนครินทรวิโรฒ ได้รับอนุมัติให้ดำเนินการทำวิทยานิพนธ์ เรื่อง “การเปรียบเทียบประสิทธิภาพของ ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในแบบทดสอบที่มีการให้คะแนนแบบหลายค่าระหว่างวิธี GMH, วิธี Polytomous SIBTEST และวิธี LDFA” ทั้งนี้ผู้วิจัยมีความต้องการเก็บข้อมูลเพื่อการวิจัย จากนักเรียนระดับชั้นมัธยมศึกษา ชั้นปีที่ 3 โรงเรียนมัธยมศึกษา สังกัดกรมสามัญศึกษา จังหวัดพระนครศรีอยุธยา

สำนักงานสามัญศึกษาจังหวัดพระนครศรีอยุธยา จึงทราบบและให้ความร่วมมือในการตอบแบบสอบถาม ดังกล่าว

จึงเรียนมาเพื่อทราบ และดำเนินการ

ขอแสดงความนับถือ

(นายนิวัตร นาคะเวช)

ผู้อำนวยการสามัญศึกษาจังหวัดพระนครศรีอยุธยา

ฝ่ายพัฒนาการศึกษา

โทร. 0-3534-5354

ภาคผนวก ง

แบบทดสอบ เรื่อง วงกลม

แบบทดสอบ เรื่อง วงกลม

คำชี้แจง

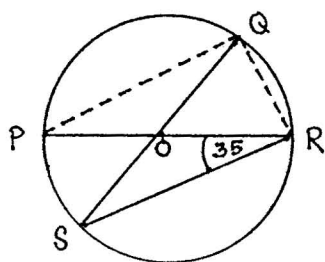
1. แบบทดสอบเรื่อง วงกลม มีข้อสอบทั้งหมด 40 ข้อ
 ตอนที่ 1 มีจำนวน 20 ข้อ (ข้อ 1 – ข้อ 20)
 ตอนที่ 2 มีจำนวน 20 ข้อ (ข้อ 21 – ข้อ 40)
2. ข้อสอบแต่ละข้อมีคำตอบมากกว่า 1 ตัวเลือก ให้นักเรียนพิจารณาตัวเลือกทุกตัว
 ถ้าเห็นว่าตัวเลือกใด ถูกต้อง ให้ทำเครื่องหมาย $\checkmark$
 ถ้าเห็นว่าตัวเลือกใด ไม่ถูกต้อง ให้ทำเครื่องหมาย X

ตัวอย่างข้อสอบ

ข้อ (A) กำหนดให้ O เป็นจุดศูนย์กลางของวงกลม A, B, C และ D เป็นจุดบนเส้นรอบวง และ $\overline{AC}$ เป็นเส้นผ่านศูนย์กลาง ถ้า $\angle ADO = 20^\circ$ แล้ว

- ก. $\angle DAO = 20^\circ$
- ข. $\angle ODC = 70^\circ$
- ค. $\angle ABC = 90^\circ$
- ง. $\angle DOC = 140^\circ$
- จ. $\angle AOD = 2(\angle ACD)$

ข้อ (B) กำหนดให้ O เป็นจุดศูนย์กลางของวงกลม และ $\angle PRS = 35^\circ$



- ก. $\angle PQO = 35^\circ$
- ข. $\angle QOR = 70^\circ$
- ค. $\angle OQR = 35^\circ$
- ง. $\angle QPO = \angle OSR$
- จ. คอร์ด PQ = คอร์ด RS

ตัวอย่างกระดาษคำตอบ

ข้อ	ก	ข	ค	ง	จ
A	$\checkmark$	$\checkmark$	$\checkmark$	X	$\checkmark$
B	$\checkmark$	$\checkmark$	X	$\checkmark$	$\checkmark$

1. ถ้าวงกลมวงหนึ่ง มีรัศมียาว 10 เซนติเมตร

ก. เส้นรอบวงยาว 20π เซนติเมตร

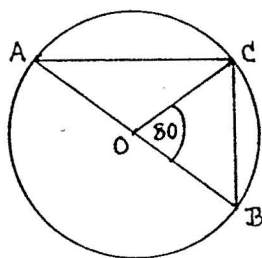
ข. คอร์ดที่ยาวที่สุดจะยาว 20 เซนติเมตร

ค. เส้นสัมผัสทุกเส้น จะอยู่ห่างจากจุดศูนย์กลาง 20 เซนติเมตร

ง. มุมในครึ่งวงกลมทุกมุมจะมีเส้นตรงที่เป็นฐานรองรับมุมยาว 10 เซนติเมตร

จ. เส้นสัมผัส 2 เส้นที่ขนานกันของวงกลมวงหนึ่ง จะอยู่ห่างกัน 20 เซนติเมตร

2. กำหนดให้ $\overline{AB}$ เป็นเส้นผ่านศูนย์กลางของวงกลม O และ $\angle BOC = 80^\circ$



ก. $\angle AOC = 90^\circ$

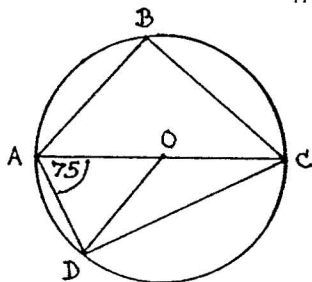
ข. $\angle OCA = 50^\circ$

ค. $\angle OCB = 40^\circ$

ง. $\angle ACB = 90^\circ$

จ. $\angle CAB = \angle CBA$

3. กำหนดให้ $\overline{AC}$ เป็นเส้นผ่านศูนย์กลางของวงกลม O และ $\angle OAD = 75^\circ$



ก. $\angle AOD = 40^\circ$

ข. $\angle DOC = 140^\circ$

ค. $\angle ODC = 20^\circ$

ง. $\angle ACD = 15^\circ$

จ. $\angle ABC = 90^\circ$

4. กำหนดให้ A, B และ C เป็นจุดบนเส้นรอบวงของวงกลม O, $\overline{AC}$ เป็นเส้นผ่านศูนย์กลาง ถ้า $\angle BAO = 60^\circ$ แล้ว

ก. $\angle AOB = 60^\circ$

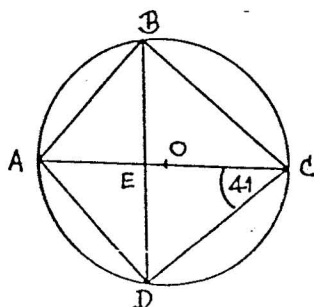
ข. $\triangle ABO$ เป็นสามเหลี่ยมด้านเท่า

ค. $\angle BOC = 100^\circ$

ง. $\angle OBC = 60^\circ$

จ. $\angle ABO + \angle OBC = 90^\circ$

5. กำหนดให้ $\overline{AC}$ เป็นเส้นผ่านศูนย์กลางของวงกลม O และมี $\overline{BD} \perp \overline{AC}$ ที่จุด E ถ้า $\angle OCD = 41^\circ$ แล้ว



ก. $\angle ABD = 41^\circ$

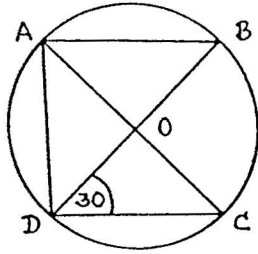
ข. $\angle ADC = \angle ABC$

ค. $\angle BAC = \angle DAC$

ง. $\angle BDC = 49^\circ$

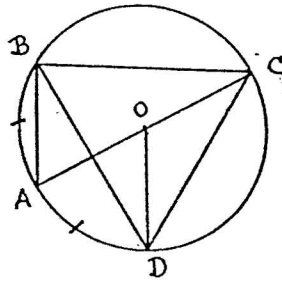
จ. $\angle AED = 2(\angle ABD)$

6. กำหนดให้ $\overline{AC}$ และ $\overline{BD}$ เป็นเส้นผ่านศูนย์กลางของวงกลม O และ $\widehat{BDC} = 30^\circ$



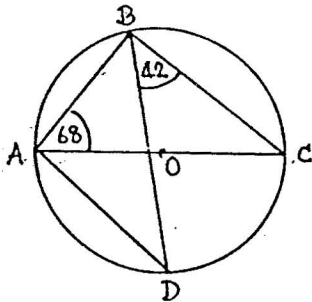
- ก. $\angle OCD = 30^\circ$
 ข. $\angle AOB = 120^\circ$
 ค. $\angle DAC = 50^\circ$
 ง. $\angle AOD = 90^\circ$
 จ. $\angle OAB = \angle OAD$

7. กำหนดให้ O เป็นจุดศูนย์กลางของวงกลม $\overline{AC}$ เป็นเส้นผ่านศูนย์กลาง ถ้า $m(\widehat{AB}) = m(\widehat{AD})$ แล้ว



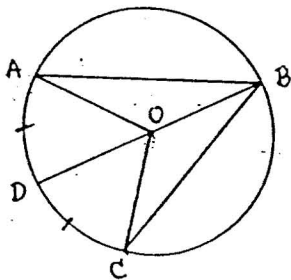
- ก. $\angle ACB = \angle ACD$
 ข. $\angle ABD + \angle ACB = \angle AOD$
 ค. $\angle OCB = \angle OCD$
 ง. $\angle BAC = \angle BDO + \angle ODC$
 จ. $\overline{BD} = \overline{BC}$

8. กำหนดให้ $\overline{AC}$ เป็นเส้นผ่านศูนย์กลางของวงกลม O ถ้า $\widehat{BAC} = 68^\circ$ และ $\widehat{DBC} = 42^\circ$ แล้ว



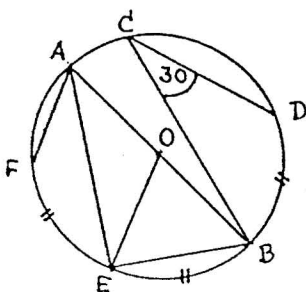
- ก. $\angle DAC = 42^\circ$
 ข. $\angle ABD = 48^\circ$
 ค. $\angle ACB = 42^\circ$
 ง. $\angle ADB = 42^\circ$
 จ. $\angle BAD = 100^\circ$

9. กำหนดให้ $\overline{DB}$ เป็นเส้นผ่านศูนย์กลางของวงกลม O ถ้า $m(\widehat{AD}) = m(\widehat{DC})$ แล้ว



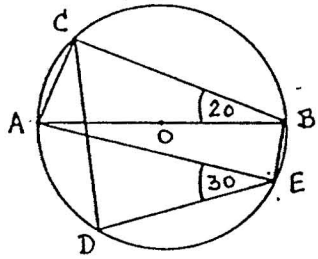
- ก. $m(\widehat{AB}) = m(\widehat{CB})$
 ข. $\angle AOB = 2(\angle COB)$
 ค. $\overline{AB} = \overline{CB}$
 ง. $\angle ABC = \angle AOD$
 จ. $\angle AOD + \angle DOC = 2(\angle ABC)$

10. กำหนดให้ $\overline{AB}$ เป็นเส้นผ่านศูนย์กลางของวงกลม O , $m(\widehat{FE}) = m(\widehat{EB}) = m(\widehat{BD})$ และ $\widehat{BCD} = 30^\circ$



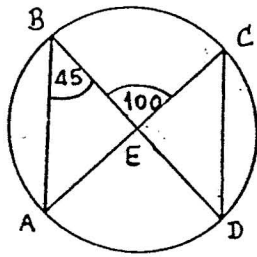
- ก. $\angle FAE = 30^\circ$
 ข. $\angle AEO = 30^\circ$
 ค. $\angle AOE = 100^\circ$
 ง. $\angle EOB = 60^\circ$
 จ. $\triangle EOB$ เป็นสามเหลี่ยมด้านเท่า

11. กำหนดให้ $\overline{AB}$ เป็นเส้นผ่านศูนย์กลางของวงกลม O , $\hat{AED} = 30^\circ$ และ $\hat{ABC} = 20^\circ$



- ก. $\hat{ACD} = 30^\circ$
 ข. $\hat{DCB} = 50^\circ$
 ค. $\hat{AEB} = \hat{BCA}$
 ง. $\hat{BAC} = 50^\circ$
 จ. $\hat{BAC} = \hat{CDE}$

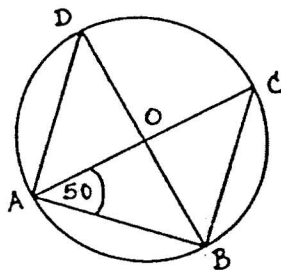
12.



- กำหนดให้ $\hat{ABE} = 45^\circ$ และ $\hat{BEC} = 100^\circ$

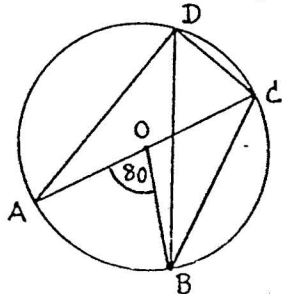
- ก. $\hat{ACD} = 45^\circ$
 ข. $\hat{DEC} = 80^\circ$
 ค. $\hat{EDC} = 50^\circ$
 ง. $\hat{BEA} = 80^\circ$
 จ. $\overline{AB} \parallel \overline{CD}$

13. กำหนดให้ $\overline{AC}$ และ $\overline{BD}$ เป็นเส้นผ่านศูนย์กลางของวงกลม O และ $\hat{OAB} = 50^\circ$



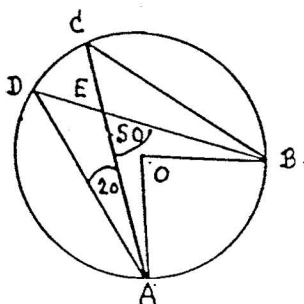
- ก. $\hat{OBA} = 50^\circ$
 ข. $\hat{ACB} = 50^\circ$
 ค. $\hat{ACB} = \hat{ADB}$
 ง. $\hat{DAC} = \hat{DBC} = 40^\circ$
 จ. $\hat{DOC} = 100^\circ$

14. กำหนดให้ $\overline{AC}$ เป็นเส้นผ่านศูนย์กลางของวงกลม O และ $\hat{AOB} = 80^\circ$



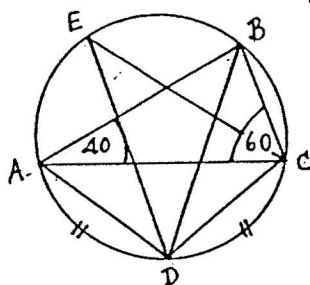
- ก. $\hat{BOC} = 100^\circ$
 ข. $\hat{BDC} = 40^\circ$
 ค. $\hat{ADB} = 50^\circ$
 ง. $\hat{ACB} = \hat{ADB}$
 จ. $\hat{DAC} = \hat{DBC}$

15. กำหนดให้ O เป็นจุดศูนย์กลางของวงกลม $\hat{AEB} = 50^\circ$ และ $\hat{DAE} = 20^\circ$



- ก. $\hat{DBC} = 20^\circ$
 ข. $\hat{DEA} = 120^\circ$
 ค. $\hat{ADE} = 40^\circ$
 ง. $\hat{ACB} = 40^\circ$
 จ. $\hat{AOB} = 100^\circ$

16.

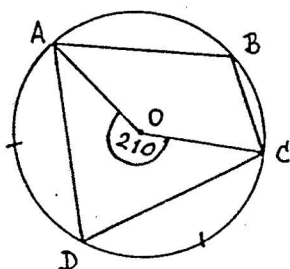
กำหนดให้ $m(\widehat{AD}) = m(\widehat{DC})$, $\widehat{BAC} = 40^\circ$ และ $\widehat{ACB} = 60^\circ$

- ก. $\widehat{ABC} = 90^\circ$
 ข. $\widehat{ABD} = \widehat{DCA}$
 ค. $\widehat{DEC} = 60^\circ$
 ง. $\widehat{ACD} = 40^\circ$
 จ. $\widehat{ADC} = 120^\circ$

17. กำหนดให้ P, Q, R และ S เป็นจุดบนเส้นรอบวงของวงกลม O โดยมี $\overline{PR}$ เป็นเส้นผ่านศูนย์กลาง $\overline{QS}$ ตัด $\overline{PR}$ ที่ T ถ้า $\widehat{QPR} = 68^\circ$ และ $\widehat{PRS} = 52^\circ$ แล้ว

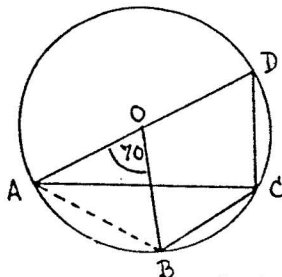
- ก. $\widehat{PQS} = 38^\circ$
 ข. $\widehat{QRP} = 52^\circ$
 ค. $\widehat{QSR} = 68^\circ$
 ง. $\widehat{OSR} = 22^\circ$
 จ. $\widehat{TSO} = 16^\circ$

18. กำหนดให้ วงกลม O มี $m(\widehat{AD}) = m(\widehat{DC})$ และ มุมกลับ $\widehat{AOC} = 210^\circ$



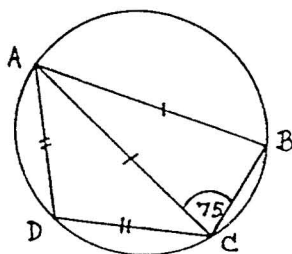
- ก. $\widehat{AOC} = 150^\circ$
 ข. $\widehat{ADC} = 105^\circ$
 ค. $\widehat{ABC} = 150^\circ$
 ง. $\widehat{AOD} = \widehat{DOC}$
 จ. $\widehat{OCD} = 75^\circ$

19. กำหนดให้ $\overline{AD}$ เป็นเส้นผ่านศูนย์กลางของวงกลม O และ $\widehat{AOB} = 70^\circ$



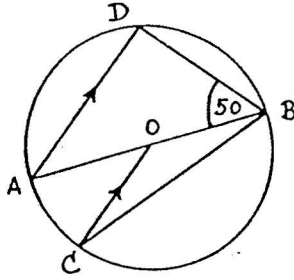
- ก. $\widehat{ACD} = 90^\circ$
 ข. $\widehat{ACB} = 20^\circ$
 ค. $\widehat{BCD} = 110^\circ$
 ง. $\widehat{BOD} = 100^\circ$
 จ. $\widehat{ABO} = 70^\circ$

20. กำหนดให้ ABCD เป็นสี่เหลี่ยมที่แนบในวงกลม $\overline{AD} = \overline{DC}$, $\overline{AC} = \overline{AB}$ และ $\widehat{ACB} = 75^\circ$



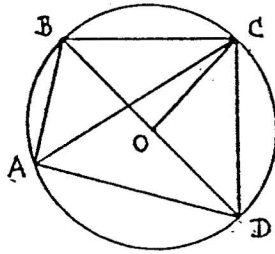
- ก. $\triangle ADC$ เป็นสามเหลี่ยมหน้าจั่ว
 ข. $\triangle ABC$ เป็นสามเหลี่ยมด้านเท่า
 ค. $\widehat{BAC} = 30^\circ$
 ง. $m(\widehat{AD}) = m(\widehat{DC})$
 จ. $\widehat{ADC} = 100^\circ$

21. กำหนดให้ $\overline{AB}$ เป็นเส้นผ่านศูนย์กลางของวงกลม O, $\overline{AD} \parallel \overline{CO}$ และ $\angle ABD = 50^\circ$



- ก. $\overline{AD} \perp \overline{BD}$
 ข. $\angle BAD = 50^\circ$
 ค. $\angle AOC = 40^\circ$
 ง. $\angle DAB = \angle OCB$
 จ. $\angle DBC = 70^\circ$

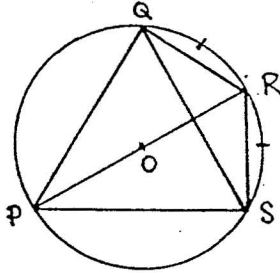
22.



กำหนดให้ $\overline{BD}$ เป็นเส้นผ่านศูนย์กลางของวงกลม O

- ก. $\angle ABC = 90^\circ$
 ข. $\angle ABD = \angle OCD$
 ค. $\angle DAC = \angle DBC$
 ง. $\angle BCO + \angle ODC = \angle DAB$
 จ. $\angle ODC + \angle BAC = \angle BOC$

23. กำหนดให้ $\overline{PR}$ เป็นเส้นผ่านศูนย์กลางของวงกลม O และ $m(\widehat{QR}) = m(\widehat{RS})$

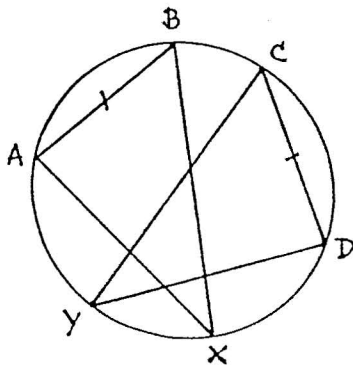


- ก. $\overline{QR} = \overline{RS}$
 ข. $\angle QRS = 90^\circ$
 ค. $\angle RPQ = \angle SQR$
 ง. $\angle QPR = \angle RPS$
 จ. $m(\widehat{PQ}) = m(\widehat{PS})$

24. กำหนดให้ $\overline{AC}$ และ $\overline{BD}$ เป็นเส้นผ่านศูนย์กลางของวงกลม O, คอร์ด AB ขนานกับคอร์ด CD

- ก. $\angle ABD = \angle BDC$
 ข. $\angle ABC = \angle BCD = 90^\circ$
 ค. $\angle ABO + \angle OCD = 90^\circ$
 ง. $\angle BAC + \angle BDC = \angle BOC$
 จ. $\angle ABO + \angle BCO = \angle OBC + \angle OCD$

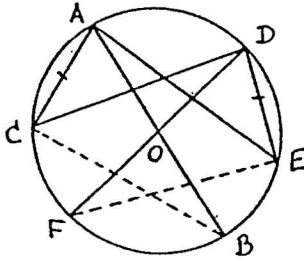
25.



กำหนดให้ $\overline{AB} = \overline{CD}$

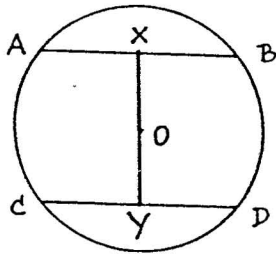
- ก. $\angle AXB = \angle CYD$
 ข. $m(\widehat{AB}) = m(\widehat{CD})$
 ค. $m(\widehat{BC}) = m(\widehat{XY})$
 ง. $\overline{BX} = \overline{CY}$
 จ. $m(\widehat{ABC}) = m(\widehat{BCD})$

26. กำหนดให้ $\overline{AB}$ และ $\overline{DF}$ เป็นเส้นผ่านศูนย์กลางของวงกลม O และ $\overline{AC} = \overline{DE}$



- ก. $m(\widehat{AC}) = m(\widehat{DE})$
 ข. $m(\widehat{CF}) = m(\widehat{BE})$
 ค. $\angle CAB = \angle FDE$
 ง. $\angle CAE = \angle CDE$
 จ. $\angle ACD = \angle DFE$

27. กำหนดให้ $\overline{AB}$ และ $\overline{CD}$ เป็นคอร์ดของวงกลม O ที่ขนานกัน $\overline{AB} = \overline{CD} = 40$ หน่วย



$\overline{XY}$ ตั้งฉากกับคอร์ดทั้งสองและ $\overline{XY}$ ยาว 30 หน่วย

- ก. $\overline{OX} = \overline{OY}$
 ข. $\overline{CY} = \overline{BX}$
 ค. $\overline{OD}$ ยาว 50 หน่วย
 ง. เส้นรอบวงยาว 100π หน่วย
 จ. $\triangle OYD \cong \triangle OYX$

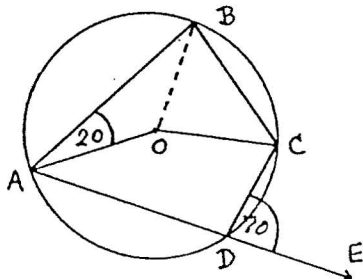
28. กำหนดให้ $\overline{AB}$ และ $\overline{CD}$ เป็นคอร์ดของวงกลม O ซึ่งอยู่ห่างจากจุดศูนย์กลาง 3 และ 4 หน่วยตามลำดับ $\overline{AB} = 8$ หน่วย

- ก. คอร์ด CD ยาวกว่าคอร์ด AB
 ข. รัศมีของวงกลม ยาว 4 หน่วย
 ค. คอร์ด CD ยาวกว่ารัศมีของวงกลม อยู่ 2 หน่วย
 ง. เส้นผ่านศูนย์กลาง ยาวกว่าคอร์ด CD 4 หน่วย
 จ. ความยาวครึ่งหนึ่งของคอร์ด CD เท่ากับระยะที่คอร์ด AB อยู่ห่างจากจุดศูนย์กลาง

29. กำหนดให้ $\overline{AB}$ เป็นเส้นผ่านศูนย์กลางของวงกลม O , $\overline{AB}$ ยาว 20 หน่วย
 คอร์ด CD อยู่ห่างจากจุดศูนย์กลาง 8 หน่วย จุด E เป็นจุดกึ่งกลางคอร์ด CD แล้ว

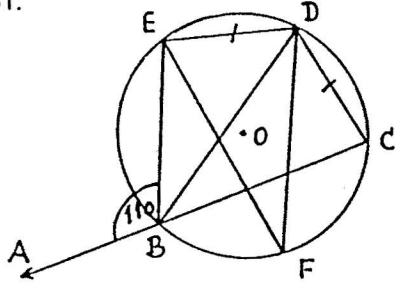
- ก. $\overline{OC} = \overline{OD} = 8$ หน่วย
 ข. $\overline{CE} = 4$ หน่วย
 ค. $\overline{CD}$ ยาวกว่ารัศมีของวงกลม อยู่ 2 หน่วย
 ง. $\overline{OE} \perp \overline{CD}$ ที่จุด E
 จ. $\triangle OCE \cong \triangle ODE$

30. กำหนดให้ O เป็นจุดศูนย์กลางของวงกลม $\angle CDE = 70^\circ$ และ $\angle OAB = 20^\circ$

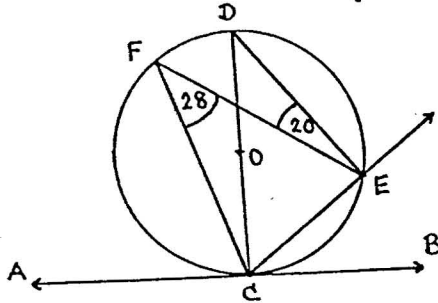


- ก. $\angle ADC = 90^\circ$
 ข. $\angle ABC = 60^\circ$
 ค. $\angle OBA = 20^\circ$
 ง. $\angle OBC = 60^\circ$
 จ. $\angle AOC = 150^\circ$

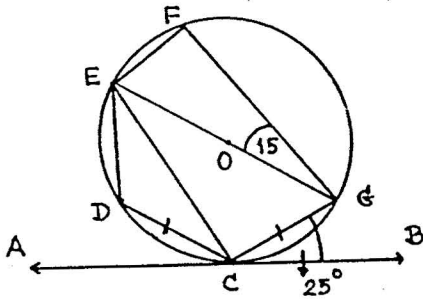
31.

กำหนดให้ $\overline{DE} = \overline{DC}$ และ $\widehat{ABE} = 110^\circ$

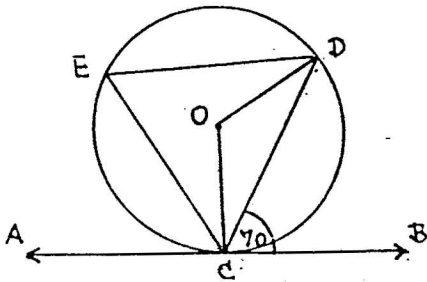
- ก. $m(\widehat{DE}) = m(\widehat{DC}) = m(\widehat{CF})$
 ข. $\widehat{DBC} = 30^\circ$
 ค. $\widehat{ABE} = \widehat{EDC}$
 ง. $\widehat{DFE} = \widehat{DBE}$
 จ. $\widehat{BEF} = \widehat{DBC}$

32. กำหนดให้ $\overleftrightarrow{CD}$ เป็นเส้นผ่านศูนย์กลาง $\overleftrightarrow{AB}$ สัมผัสวงกลม O ที่จุด C และ $\widehat{CFE} = 28^\circ$, $\widehat{FED} = 20^\circ$ 

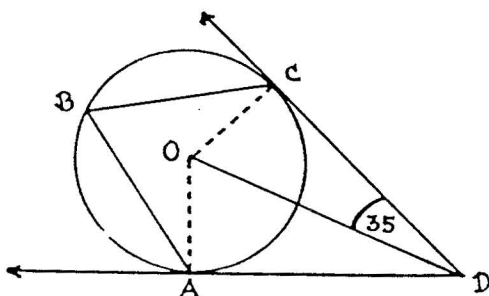
- ก. $\widehat{CDE} = 28^\circ$
 ข. $\widehat{FCD} = 28^\circ$
 ค. $\widehat{FEC} = 62^\circ$
 ง. $\widehat{BCE} = \widehat{CDE}$
 จ. $\widehat{ACF} = 62^\circ$

33. กำหนดให้ $\overleftrightarrow{AB}$ สัมผัสวงกลม O ที่จุด C, $\overline{DC} = \overline{CG}$, $\widehat{BCG} = 25^\circ$ และ $\widehat{EGF} = 15^\circ$ 

- ก. $\widehat{CEG} = 25^\circ$
 ข. $\widehat{CGE} = 75^\circ$
 ค. $\widehat{FEG} = 75^\circ$
 ง. $\widehat{DEG} = 60^\circ$
 จ. $\widehat{ACD} = 15^\circ$

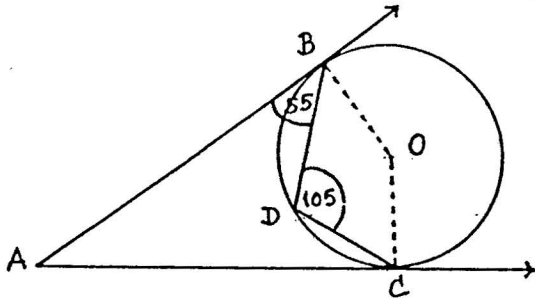
34. กำหนดให้ $\overleftrightarrow{AB}$ เป็นเส้นสัมผัส ซึ่งสัมผัสวงกลม O ที่จุด C, $\widehat{BCD} = 70^\circ$ 

- ก. $\widehat{ODC} = 30^\circ$
 ข. $\widehat{COD} = 110^\circ$
 ค. $\widehat{ODC} = \widehat{OCD}$
 ง. $\widehat{CED} = \widehat{ODC} + \widehat{OCD}$
 จ. มุมกลับ $\widehat{COD} = 220^\circ$

35. กำหนดให้ $\overrightarrow{DA}$ และ $\overrightarrow{DC}$ สัมผัสวงกลม O ที่จุด A และ C ตามลำดับ และ $\widehat{ODC} = 35^\circ$ 

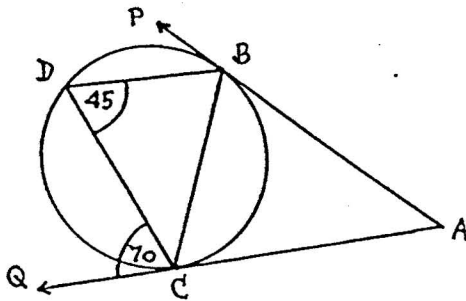
- ก. $\widehat{ADO} = 35^\circ$
 ข. $\overline{AD} = \overline{DC}$
 ค. $\triangle AOD \cong \triangle COD$
 ง. $\widehat{AOD} = 50^\circ$
 จ. $\widehat{ABC} = 100^\circ$

36. กำหนดให้ $\vec{AB}$ และ $\vec{AC}$ สัมผัสวงกลม O ที่จุด B และ C ตามลำดับ และ $\hat{ABD} = 55^\circ$, $\hat{BDC} = 105^\circ$



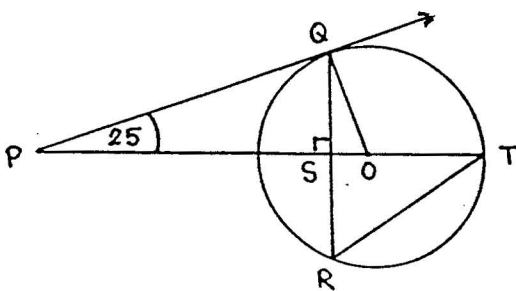
- ก. $\hat{OBD} = 35^\circ$
 ข. $\hat{BOD} = 100^\circ$
 ค. $\hat{ODC} = 65^\circ$
 ง. $\hat{ACD} = 25^\circ$
 จ. $\hat{BOC} = 155^\circ$

37. กำหนดให้ $\vec{AP}$ และ $\vec{AQ}$ สัมผัสวงกลมที่ B และ C ตามลำดับ $\hat{BDC} = 45^\circ$ และ $\hat{DCQ} = 70^\circ$



- ก. $\hat{DBC} = 65^\circ$
 ข. $\hat{BCD} = 70^\circ$
 ค. $\hat{ACB} = 60^\circ$
 ง. $\hat{DBP} = 65^\circ$
 จ. $\hat{BAC} = 80^\circ$

38. กำหนดให้ $\vec{PQ}$ สัมผัสวงกลม O ที่จุด Q , คอร์ด QR ตั้งฉากกับเส้นผ่านศูนย์กลางที่จุด S , $\hat{QPO} = 25^\circ$



- ก. $\hat{PQO} = 90^\circ$
 ข. $\hat{PQS} = 55^\circ$
 ค. $\hat{QOS} = 55^\circ$
 ง. $\hat{QOT} = 120^\circ$
 จ. $\hat{QRT} = 60^\circ$

39. ก. มุมที่จุดศูนย์กลางของรูป 12 เหลี่ยมด้านเท่ามุมเท่า มีขนาดมุมละ 30 องศา
 ข. รูป 20 เหลี่ยมด้านเท่ามุมเท่า มีมุมภายในแต่ละมุมมีขนาด 160 องศา
 ค. มุมภายในแต่ละมุมของรูป 4 เหลี่ยมด้านเท่ามุมเท่า เป็นมุมฉาก
 ง. รูปเหลี่ยมที่มีมุมภายในแต่ละมุมมีขนาด 168 องศา เป็นรูป 30 เหลี่ยมด้านเท่ามุมเท่า
 จ. ผลบวกของขนาดของมุมภายในทุกมุมของรูป 15 เหลี่ยมด้านเท่ามุมเท่า เท่ากับ 2300 องศา

40. กำหนดให้ ABCDEFGHIJ เป็นรูป 10 เหลี่ยมด้านเท่ามุมเท่า แนบในวงกลม O

- ก. $\hat{AOB} = 30^\circ$
 ข. $\hat{ABC} = 144^\circ$
 ค. $\hat{AOE} = 144^\circ$
 ง. $\overline{AF}$ เป็นเส้นผ่านศูนย์กลาง
 จ. ผลบวกของขนาดของมุมภายในทุกมุม เท่ากับ 1440°

ภาคผนวก จ

ตาราง 29 ผลการประเมินความสอดคล้องระหว่างเนื้อหา กับระดับพฤติกรรมที่ต้องการวัด และการพิจารณา
การทำหน้าที่ต่างกันของข้อสอบ

ระดับพฤติกรรม*	ข้อสอบที่สร้าง	ผู้เชี่ยวชาญคนที่								รวม	IOC 2	แบบทดสอบฉบับ 1 (50 ข้อ)	คุณภาพข้อสอบ (Tryout I)		แบบทดสอบฉบับ 2 (40 ข้อ)	คุณภาพข้อสอบ (Tryout II)		ขนาดแบบทดสอบ (ข้อ)		
		IOC 1 (DIF)											p	r		p	r	40	30	20
		1	2	3	4	5														
1	(1)	1	0	1	1	1	1	4	0.8	(1)	0.68	0.39	(1)	0.64	0.36	(1)	(1)	(1)		
	(2)	1	0	1	1	1	1	4	0.8	-	-	-	-	-	-	-	-			
	(3)	1	0	1	1	1	1	4	0.8	-	-	-	-	-	-	-	-			
	(4)	1	1	1	1	1	1	5	1	-	-	-	-	-	-	-	-			
	(5)	1	1	1	1	1	1	5	1	-	-	-	-	-	-	-	-			
	(6)	1	1	1	1	1	1	5	1	-	-	-	-	-	-	-	-			
	(7)	1	1	0	1	1	1	4	0.8	-	-	-	-	-	-	-	-			
	(8)	1	1	1	1	1	1	5	1	(2)	0.70	0.30	(39)	0.64	0.20	(39)	(30)	(20)		
2	9	1	0	1	1	1	1	4	0.8	3	0.75	0.50	2	0.69	0.27	2	2	2		
	10	1	0	1	1	1	1	4	0.8	4	0.64	0.43	3	0.69	0.61	3	-	3		
	12	1	0	1	1	1	1	4	0.8	6	0.52	0.38	5	0.68	0.30	5	-	-		
	13	1	0	1	1	1	1	4	0.8	7	0.64	0.10	-	-	-	-	-	-		
	14	1	1	1	1	1	1	5	1	8	0.66	0.29	6	0.54	0.25	6	3	4		
	15	1	0	1	1	1	1	4	0.8	9	0.73	0.26	7	0.55	0.39	7	-	-		
	16	1	0	1	1	1	1	4	0.8	10	0.58	0.39	8	0.64	0.37	8	4	5		
	19	1	0	1	1	1	1	4	0.8	13	0.59	0.39	11	0.80	0.32	11	6	6		
	20	1	0	1	1	1	1	4	0.8	14	0.57	0.37	12	0.57	0.45	12	7	-		
	21	1	0	1	1	1	1	4	0.8	15	0.71	0.44	13	0.54	0.30	13	8	7		
	22	1	0	1	1	1	1	4	0.8	16	0.80	0.47	14	0.68	0.47	14	9	8		
	23	1	0	0	1	1	1	3	0.6	17	0.70	0.41	15	0.73	0.51	15	-	9		
	24	1	0	1	1	1	1	4	0.8	18	0.69	0.31	16	0.60	0.46	16	10	-		
	25	1	0	1	1	1	1	4	0.8	19	0.63	0.20	-	-	-	-	-	-		
	(26)	1	0	1	1	1	1	4	0.8	(20)	0.78	0.13	-	-	-	-	-	-		
	(27)	1	0	1	1	1	1	4	0.8	-	-	-	-	-	-	-	-	-		
	30	1	0	1	1	1	1	4	0.8	23	0.51	0.52	19	0.61	0.36	19	13	11		
	38	1	0	0	1	1	1	3	0.6	30	0.67	0.33	25	0.70	0.34	25	18	-		
	39	1	0	1	1	1	1	4	0.8	31	0.91	0.14	-	-	-	-	-	-		
	40	1	0	1	1	1	1	4	0.8	32	0.80	0.44	26	0.64	0.36	26	19	-		
	45	1	0	1	1	1	1	4	0.8	37	0.54	0.35	31	0.57	0.60	31	-	-		
	46	1	0	1	1	1	1	4	0.8	38	0.67	0.57	32	0.54	0.34	32	23	16		
	47	1	0	1	1	1	1	4	0.8	39	0.59	0.33	33	0.66	0.47	33	24	-		
	49	1	0	1	1	1	1	4	0.8	41	0.73	0.45	34	0.55	0.32	34	25	17		
	(58)	1	0	0	0	1	1	2	0.4	(48)	0.61	0.10	-	-	-	-	-	-		

ตาราง 29 (ต่อ)

ระดับพฤติกรรม*	ข้อสอบที่สร้าง	IOC 1 (DIF)	ผู้เชี่ยวชาญคนที่					รวม IOC 2	แบบทดสอบฉบับ 1 (50 ข้อ)	คุณภาพข้อสอบ (Tryout I)		แบบทดสอบฉบับ 2 (40 ข้อ)	คุณภาพข้อสอบ (Tryout II)		ขนาดแบบทดสอบ (ข้อ)			
										**1			**2					
			1	2	3	4	5			p	r		p	r	40	30	20	
3	(11)	1	0	1	1	1	1	4	0.8	(5)	0.66	0.40	(4)	0.62	0.46	(4)	-	-
	17	1	1	0	0	1	1	3	0.6	11	0.80	0.50	9	0.56	0.43	9	-	-
	18	1	0	0	1	1	1	3	0.6	12	0.80	0.39	10	0.74	0.37	10	5	-
	(28)	1	0	1	1	1	1	4	0.8	(21)	0.54	0.27	(17)	0.59	0.37	(17)	(11)	-
	29	1	0	1	1	1	1	4	0.8	22	0.64	0.40	18	0.52	0.20	18	12	10
	31	1	0	1	1	1	1	4	0.8	24	0.74	0.59	20	0.50	0.48	20	14	12
	32	1	0	1	1	1	1	4	0.8	25	0.76	0.46	21	0.75	0.49	21	15	-
	(33)	1	0	0	1	1	1	3	0.6	(26)	0.70	0.18	-	-	-	-	-	-
	34	1	0	1	1	1	1	4	0.8	27	0.63	0.46	22	0.73	0.42	22	16	-
	35	1	0	1	1	1	1	4	0.8	28	0.59	0.41	23	0.64	0.32	23	17	13
	(36)	1	1	1	1	1	1	5	1	(29)	0.69	0.39	(24)	0.56	0.31	(24)	-	-
	(42)	1	1	1	1	1	1	5	1	(34)	0.59	0.34	(28)	0.80	0.30	(28)	(21)	-
	(43)	1	1	1	1	1	1	5	1	(35)	0.58	0.43	(29)	0.55	0.26	(29)	-	-
	48	1	1	0	1	1	1	4	0.8	40	0.56	0.24	-	-	-	-	-	-
	(56)	1	-1	1	1	1	1	3	0.6	-	-	-	-	-	-	-	-	-
	57	1	0	1	1	1	1	4	0.8	47	0.55	0.29	37	0.56	0.38	37	28	19
	(59)	1	0	1	1	1	1	4	0.8	(49)	0.62	0.25	(40)	0.55	0.30	(40)	-	-
	60	1	1	1	1	1	1	5	1	50	0.51	0.37	38	0.52	0.33	38	29	-
4	37	1	0	1	1	1	1	4	0.8	-	-	-	-	-	-	-	-	-
	41	1	0	1	1	1	1	4	0.8	33	0.80	0.27	27	0.80	0.40	27	20	14
	44	1	0	1	0	1	1	3	0.6	36	0.65	0.52	30	0.56	0.30	30	22	15
	(50)	1	0	1	1	1	1	4	0.8	(42)	0.67	0.25	-	-	-	-	-	-
	(51)	1	0	1	1	1	1	4	0.8	(43)	0.49	0.13	-	-	-	-	-	-
	52	1	1	1	1	1	1	5	1	44	0.67	0.10	-	-	-	-	-	-
	53	1	0	1	1	1	1	4	0.8	-	-	-	-	-	-	-	-	-
	54	1	0	1	1	1	1	4	0.8	45	0.80	0.40	35	0.63	0.57	35	26	18
	55	1	0	1	1	1	1	4	0.8	46	0.59	0.28	36	0.80	0.48	36	27	-

หมายเหตุ IOC1 หมายถึง ผลการพิจารณาว่าข้อสอบแสดงการทำหน้าที่ต่างกัน (DIF)

IOC2 หมายถึง ผลการประเมินความสอดคล้องระหว่างเนื้อหากับระดับพฤติกรรมที่ต้องการวัด

**1 แทน ค่าความเชื่อมั่นของแบบทดสอบฉบับ 1 (α) เท่ากับ 0.8839

**2 แทน ค่าความเชื่อมั่นของแบบทดสอบฉบับ 2 (α) เท่ากับ 0.8846

ข้อสอบที่อยู่ในวงเล็บ () หมายถึง ข้อสอบที่สร้างให้แสดงการทำหน้าที่ต่างกัน (DIF)

ระดับพฤติกรรม* หมายถึง ระดับพฤติกรรมการเรียนรู้ 4 ชั้น ตามแนวคิดของวิลสัน (Wilson)

ตาราง 31 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) เมื่อกลุ่มตัวอย่างมีขนาด 500 คน และแบบทดสอบมีขนาด 20 ข้อ

กลุ่มที่	วิธีการตรวจสอบ	ข้อสอบที่แสดงการทำหน้าที่ต่างกัน (DIF)													รวม	ผล	
	GPCM	1	2	3	6	7	10	13	14	15	17	18	20	12	P	E	
1	GMH	13	14	20										3	25	0	
	Poly-SIBTEST	2	7	9	12	13	14	18	20					8	50	25	
2	GMH	10	13	20										3	25	0	
	Poly-SIBTEST	2	9	10	12	13	14	15	17	18	20			10	66.67	25	
3	GMH	5	9	12	18									4	8.33	37.5	
	Poly-SIBTEST	5	9	11	12	13	14	15	16	17	18			10	41.67	62.5	
4	GMH	7	11	13	14	19	20							6	33.33	25	
	Poly-SIBTEST	2	7	9	11	12	13	14	20					8	41.67	37.5	
5	GMH	12	14	18										3	16.67	12.5	
	Poly-SIBTEST	2	7	8	11	12	13	14	15	18	20			10	58.33	37.5	
6	GMH	12	13	15	17	19								5	25	25	
	Poly-SIBTEST	2	12	13	14	18	20							6	41.67	12.5	
7	GMH	3	7	8	12	13	17							6	33.33	25	
	Poly-SIBTEST	2	7	8	9	12	13	15	20					8	41.67	37.5	
8	GMH	9	12	20										3	8.33	25	
	Poly-SIBTEST	2	8	9	12	14	15	18						7	33.33	37.5	
9	GMH	2	7	17										3	25	0	
	Poly-SIBTEST	2	7	9	14	16	18	20						7	41.67	25	
10	GMH	4	7	12	13	18								5	25	25	
	Poly-SIBTEST	2	7	8	9	10	12	13	14	15	18	19	20	12	66.67	50	
														GMH	22.5	15.83	
														X	Poly-SIBTEST	48.34	35.0
														GMH	8.83	13.44	
														SD	Poly-SIBTEST	11.65	14.19

ตาราง 32 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) เมื่อกลุ่มตัวอย่างมีขนาด 250 คน และแบบทดสอบมีขนาด 20 ข้อ

กลุ่มที่	วิธีการตรวจสอบ	ข้อสอบที่แสดงการทำหน้าที่ต่างกัน (DIF)	รวม	ผล	
	GPCM	1 2 3 6 7 10 13 14 15 17 18 20		P	E
1	GMH	5 14	2	8.33	12.5
	Poly-SIBTEST	10 13 19	3	16.67	12.5
2	GMH	12 14	2	8.33	12.5
	Poly-SIBTEST	5 12 13 14 16 18	6	25	37.5
3	GMH	2 13	2	16.67	0
	Poly-SIBTEST	2 6 15	3	25	0
4	GMH	4 9	2	0	25
	Poly-SIBTEST	9 12 14 16	4	8.33	37.5
5	GMH	7 16 18	3	16.67	12.5
	Poly-SIBTEST	2 7 9 15 16	5	25	25
6	GMH	-	0	0	0
	Poly-SIBTEST	1 2 7 12 13 14	6	41.67	12.5
7	GMH		0	0	0
	Poly-SIBTEST	6 12	2	8.33	12.5
8	GMH	3 12 13	3	16.67	12.5
	Poly-SIBTEST	7 9 11 12 18	5	16.67	37.5
9	GMH	13	1	8.33	0
	Poly-SIBTEST	1 2 13 18 20	5	41.67	0
10	GMH	10 13 14 18	4	33.33	0
	Poly-SIBTEST	1 2 6 10 13 14 15 19 20	9	66.67	12.5
			GMH	10.83	7.50
			$\bar{X}$ Poly-SIBTEST	27.50	18.75
			GMH	10.43	8.74
			SD Poly-SIBTEST	18.03	14.73

ตาราง 33 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) เมื่อกลุ่มตัวอย่างมีขนาด 1,000 คน และแบบทดสอบมีขนาด 30 ข้อ

กลุ่มที่	วิธีการตรวจสอบ	ข้อสอบที่แสดงการทำหน้าที่ต่างกัน (DIF)	รวม	ผล	
	GPCM	14 17 18 25 26 29		P	E
1	GMH	7 8 12 13 14 15 16 17 18 19 20 21 26	13	50	37.5
	Poly-SIBTEST	2 5 7 8 12 13 14 15 16 17 19 20 21 22 23 26 30	17	50	58.3
2	GMH	2 7 14 16 17 18 19 20 24 26	10	66.67	25
	Poly-SIBTEST	2 5 7 8 11 12 14 15 17 18 19 20 21 22 23 24 26 27 30	19	66.67	62.5
3	GMH	7 8 13 14 15 16 17 18 19 20 26 28 30	13	66.67	37.5
	Poly-SIBTEST	2 7 8 11 14 15 17 18 19 20 22 23 26 27 30	15	66.67	45.83
4	GMH	7 8 10 12 13 14 15 16 17 18 19 20 24 26 28 30	16	66.67	50
	Poly-SIBTEST	1 2 5 7 8 12 13 14 15 17 19 20 21 22 23 24 25 26 30	19	66.67	62.5
5	GMH	7 8 11 12 14 16 17 18 19 20 26 30	12	66.67	33.33
	Poly-SIBTEST	2 5 7 8 11 12 14 15 17 18 19 20 21 22 23 26 30	17	66.67	54.17
6	GMH	7 11 14 15 16 17 18 19 20 24 26 28 29 30	14	83.33	37.5
	Poly-SIBTEST	2 5 7 8 11 12 13 14 15 17 18 19 20 22 24 26 30	17	66.67	54.17
7	GMH	7 8 12 13 14 15 17 18 19 20 24 26 30	13	66.67	37.5
	Poly-SIBTEST	1 2 7 8 12 13 14 15 17 18 19 20 21 22 23 24 26 30	18	66.67	58.33
8	GMH	7 8 11 12 14 15 16 17 18 20 21 24 26 28	14	66.67	41.67
	Poly-SIBTEST	2 5 7 8 12 13 14 15 17 18 19 20 21 22 23 24 26 27 30	19	66.67	62.5
9	GMH	7 14 15 16 17 18 19 20 21 26 28 30	12	66.67	33.33
	Poly-SIBTEST	1 2 5 8 12 14 15 17 19 20 21 23 26 30	14	50	45.83
10	GMH	10 12 14 15 17 18 19 20 21 26 30	11	66.67	29.17
	Poly-SIBTEST	1 2 7 8 12 13 14 15 17 18 19 20 21 22 23 24 26 27 30	19	66.67	62.5
			GMH	66.67	36.25
			$\bar{X}$ Poly-SIBTEST	63.34	56.67
			GMH	7.86	6.82
			SD Poly-SIBTEST	7.03	6.57

ตาราง 34 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) เมื่อกลุ่มตัวอย่างมีขนาด 500 คน และแบบทดสอบมีขนาด 30 ข้อ

กลุ่มที่	วิธีการตรวจสอบ	ข้อสอบที่แสดงการทำหน้าที่ต่างกัน (DIF)	รวม	ผล	
	GPCM	14 17 18 25 26 29	6	P	E
1	GMH	16 17 20 28	4	16.67	12.5
	Poly-SIBTEST	2 14 17 20 26 28 30	7	50	16.67
	GMH	11 15 16 17	4	16.67	12.5
2	Poly-SIBTEST	5 6 8 13 14 17 18 19 20 21 26 27*30	13	66.67	37.5
	GMH	17 18 20 26	4	50	4.17
3	Poly-SIBTEST	2 7 14 15 17 19 20 21 26 30	10	50	29.17
	GMH	11 17 20 26	4	33.33	8.33
4	Poly-SIBTEST	2 8 11 12 14 15 17 19 20 23 30	11	33.33	37.5
	GMH	2 7 8 15 18 19 20 26	8	33.33	25
5	Poly-SIBTEST	2 7 8 17 18 19 20 22	8	33.33	25
	GMH	7 8 10 11 12 14 18 19	8	33.33	25
6	Poly-SIBTEST	2 5 8 10 11 12 14 16 17 19 20 24	12	33.33	41.67
	GMH	4 7 12 14 17 20 30	7	33.33	20.83
7	Poly-SIBTEST	8 14 15 17 19 20 22 24 30	9	33.33	29.17
	GMH	8 13 16 17 18 21 22 23 24 26	10	50	29.17
8	Poly-SIBTEST	2 8 14 17 18 19 20 21 23 24	10	50	29.17
	GMH	2 8 14 17 18 20 23 24 26	9	66.67	20.83
9	Poly-SIBTEST	2 5 7 8 14 17 20 23 24 26 27 30	12	50	37.5
	GMH	17 18 20 26	4	50	4.17
10	Poly-SIBTEST	2 8 14 17 18 19 20 26	8	66.67	16.67
	GMH			38.33	16.25
X	Poly-SIBTEST			46.67	30.00
	GMH			15.81	9.10
SD	Poly-SIBTEST			13.15	8.74
	GMH				

ตาราง 36 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) เมื่อกลุ่มตัวอย่างมีขนาด 1000 คน และแบบทดสอบมีขนาด 40 ข้อ

กลุ่มที่	วิธีการตรวจสอบ		ข้อสอบที่แสดงการทำหน้าที่ต่างกัน (DIF)	รวม	ผล	
	GPCM				P	E
1	GPCM	4 5 9 30 33 35 37 39		8		
	GMH	2 4 9 17 18 20 21 22 23 25 27 35 39		13	50	28.13
	Poly-SIBTEST	1 2 9 12 13 15 17 18 20 21 22 23 26 27 28 33 39 40		18	37.5	46.88
2	GMH	4 7 9 13 18 19 20 21 22 23 25 27 38 39		14	50	31.25
	Poly-SIBTEST	1 2 7 9 12 13 14 15 17 18 20 21 23 25 26 27 28 30 32 33 37 39 40		23	62.5	56.25
3	GMH	2 4 7 9 12 13 18 20 21 22 23 25 26 27 35 39		16	50	37.5
	Poly-SIBTEST	2 9 13 15 17 18 19 20 21 23 25 26 27 28 30 32 33 35 39 40		20	62.5	46.88
4	GMH	2 4 7 9 12 13 18 20 21 22 23 25 27 31 35 39		16	50	37.5
	Poly-SIBTEST	2 9 12 13 15 17 18 20 21 23 25 26 27 30 35 39 40		17	50	40.63
5	GMH	2 4 5 9 12 13 18 19 21 23 25 27 31 35		14	50	31.25
	Poly-SIBTEST	2 9 12 15 17 18 20 21 23 25 26 27 28 30 32 39 40		17	37.5	43.75
6	GMH	2 4 5 9 12 18 19 20 21 22 23 25 26 27 35 37 39 40		18	75	37.5
	Poly-SIBTEST	1 2 9 12 13 15 18 20 21 23 26 27 28 30 33 35 39 40		18	62.5	40.63
7	GMH	2 4 5 9 12 13 18 21 22 23 25 27 35 39		14	62.5	28.13
	Poly-SIBTEST	2 9 12 13 15 17 19 20 21 23 25 26 27 28 30 32 33 35 39 40		20	62.5	46.88
8	GMH	2 4 5 7 9 12 18 20 21 22 23 25 27 28 35 37 39		17	75	34.38
	Poly-SIBTEST	1 2 9 13 15 17 18 20 21 23 26 27 28 30 32 33 35 39 40		19	62.5	43.75
9	GMH	2 4 6 9 18 19 20 21 23 25 35 39		12	50	25
	Poly-SIBTEST	1 2 9 12 13 14 15 17 18 19 20 21 23 26 27 28 32 35 37 39 40		21	50	53.13
10	GMH	4 5 9 13 18 20 21 22 23 25 26 27 35 39		14	62.5	28.13
	Poly-SIBTEST	1 2 9 12 13 17 18 20 21 23 25 26 27 28 35 39 40		17	37.5	43.75
				GMH	57.5	31.88
				$\bar{X}$ Poly-SIBTEST	52.5	46.25
				GMH	10.54	4.61
				SD Poly-SIBTEST	11.49	5.06

ตาราง 37 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) เมื่อกลุ่มตัวอย่างมีขนาด 500 คน และแบบทดสอบมีขนาด 40 ข้อ

กลุ่มที่	วิธีการตรวจสอบ		ข้อสอบที่แสดงการทำหน้าที่ต่างกัน (DIF)		รวม	ผล	
	GPCM		4 5 9 30 33 35 37 39			P	E
1	GMH		10 12 20 21 23 25 35		7	12.5	18.75
	Poly-SIBTEST		1 2 9 11 12 13 18 20 21 23 27 39 40		13	25	34.38
2	GMH		13 23 25 26 28 39		6	12.5	15.63
	Poly-SIBTEST		2 7 13 18 20 21 23 26 30 32 34 37 39 40		14	37.5	34.38
3	GMH		2 4 9 18 22 38		6	25	12.5
	Poly-SIBTEST		1 2 18 20 21 23 25 27 33		9	12.5	25
4	GMH		3 5 10 13 20 22 26 28 35		9	25	21.88
	Poly-SIBTEST		2 4 19 20 23 26 27 28 35 37 38 39		12	50	25
5	GMH		2 4 12 23 26 35 37		7	37.5	12.5
	Poly-SIBTEST		2 18 20 21 22 23 24 25 26 27 28 32 39 40		14	12.5	40.63
6	GMH		2 4 18 21 23 25 27 32 40		9	12.5	25
	Poly-SIBTEST		2 3 13 18 20 21 23 25 26 27 30 32 40		13	12.5	37.5
7	GMH		4 11 12 18 21 23 27 32 39		9	25	21.88
	Poly-SIBTEST		2 13 18 20 21 23 25 26 27 32 38 39 40		13	12.5	34.38
8	GMH		5 21 23 25 35		5	25	9.38
	Poly-SIBTEST		2 5 9 13 15 20 21 22 23 27		10	25	25
9	GMH		2 4 7 9 18 25 35 37		8	50	12.5
	Poly-SIBTEST		2 9 12 18 21 23 27 32 33 37		10	37.5	21.88
10	GMH		9 12 16 23 25 26 35 37		8	37.5	15.63
	Poly-SIBTEST		1 2 9 13 15 18 20 21 23 25 27 28 30 35 36 37 39		17	62.5	37.5
					GMH	26.25	16.57
					X Poly-SIBTEST	28.75	31.57
					GMH	12.43	5.11
					SD Poly-SIBTEST	17.73	6.68

ตาราง 38 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) เมื่อกลุ่มตัวอย่างมีขนาด 250 คน และแบบทดสอบมีขนาด 40 ข้อ

กลุ่มที่	วิธีการตรวจสอบ	ข้อสอบที่แสดงการทำหน้าที่ต่างกัน (DIF)	รวม	ผล	
	GPCM	4 5 9 30 33 35 37 39		P	E
1	GMH	17	1	0	3.13
	Poly-SIBTEST	2 15 21 25 36 39 40	7	12.5	18.75
	GMH	5 26 38	3	12.5	6.25
2	Poly-SIBTEST	2 20 21 26 28 32 38	7	0	21.88
	GMH	-	0	0	0
3	Poly-SIBTEST	2 13 15 21 23 28	6	0	18.75
	GMH	13 20 21 23 31 35	6	12.5	15.63
4	Poly-SIBTEST	10 23 24 26 32 33 40	7	12.5	18.75
	GMH	5 17 23 25 35	5	25	9.38
5	Poly-SIBTEST	9 13 17 23 37 40	6	25	15.63
	GMH	18 27 28	3	0	9.38
6	Poly-SIBTEST	2 16 18 20 21 23 32 34 40	9	0	28.13
	GMH	12 18 23 35	4	12.5	9.38
7	Poly-SIBTEST	2 4 9 10 18 19 23 32	8	25	18.75
	GMH	5 9 18 21 24 32 33 35	8	50	12.5
8	Poly-SIBTEST	3 8 9 20 21 23 32 34 36	9	12.5	25
	GMH	4 11 12 17 27	5	12.5	12.5
9	Poly-SIBTEST	2 3 5 18 20 23 26 27 34 38 39 40	12	25	31.25
	GMH	4	1	12.5	0
10	Poly-SIBTEST	10 12 18 19 23 27 28 37	8	12.5	21.88
	GMH			13.75	7.82
	$\bar{X}$	Poly-SIBTEST		12.5	21.88
	SD	GMH		14.97	5.36
		Poly-SIBTEST		10.21	4.89

ภาคผนวก ข

ตัวอย่างผลการวิเคราะห์ DIF
ด้วยวิธี Polytomous SIBTEST

name of input parameter file = SIB.IN
 number of items on test = 20
 name of file for Ref. grp. scores = a:\ma20.dat
 name of file for Focal grp. scores = a:\fe20.dat
 minimum no. of examinees per matching score cell = 2
 number of runs for this data set = 20
 number of examinees in Reference Group = 1000
 number of examinees in Focal group = 1000

Examinee Test Score Summary Statistics

Reference Group: Mean = 57.28
 Standard deviation = 9.85
 Focal Group: Mean = 59.81
 Standard deviation = 11.38

Standardized Score Difference = -.24

Item Statistics

= item number
 m = mean score on item
 r = point biserial (item score-test score correlation)

#:	1	2	3	4	5	6	7	8	9	10
m:	2.696	3.086	2.953	2.616	2.569	2.767	3.165	3.237	2.665	2.897
r:	0.452	0.568	0.524	0.462	0.421	0.425	0.454	0.480	0.455	0.458
#:	11	12	13	14	15	16	17	18	19	20
m:	2.490	3.148	3.157	3.750	2.636	2.807	2.806	3.604	2.537	2.961
r:	0.489	0.481	0.396	0.420	0.548	0.461	0.494	0.453	0.391	0.275

p-value notation:

R denotes p-value for test of DIF/DBF against Ref. group
 F denotes p-value for test of DIF/DBF against Foc. group
 E denotes p-value for test of DIF/DBF against either the
 Ref. or Foc. group.

NOTES:

MS/SSD = Matching Subtest Standardized Score Difference.
 Standardized difference in mean observed scores
 between Reference group and Focal group on the
 matching subtest.

p-elim = proportion of Reference (R) and Focal (F) groups
 eliminated (not used) in SIBTEST calculations.

Positive Beta estimate indicates DIF/DBF favoring Ref. grp.
 Negative Beta estimate indicates DIF/DBF favoring Foc. grp.

FLAG = error flag indicator. FLAG=0 indicates a normal
 successful completion of a SIBTEST run. All other values
 of FLAG come with short error messages.

SIBTEST-pooled weighting

Run	Suspect	Subtest	Beta	standard	p-value	p-elim	MS		F
no.	Item	Numbers	estimate	error		R	F	SSD	L
									A
									G
1	1		0.119	0.050	0.018 E	.02	.05	-0.25	0
2	2		0.196	0.050	0.000 E	.02	.02	-0.25	0
3	3		0.008	0.054	0.888 E	.02	.03	-0.24	0
4	4		0.037	0.051	0.470 E	.01	.02	-0.24	0
5	5		0.031	0.045	0.488 E	.02	.03	-0.24	0
6	6		-0.087	0.049	0.080 E	.02	.03	-0.23	0
7	7		-0.126	0.044	0.004 E	.02	.03	-0.22	0
8	8		-0.093	0.052	0.072 E	.02	.03	-0.23	0
9	9		0.182	0.052	0.000 E	.02	.04	-0.25	0
10	10		0.132	0.050	0.008 E	.02	.02	-0.25	0
11	11		0.107	0.046	0.021 E	.02	.03	-0.25	0
12	12		-0.193	0.048	0.000 E	.02	.02	-0.22	0
13	13		0.240	0.049	0.000 E	.01	.02	-0.26	0
14	14		-0.263	0.048	0.000 E	.02	.02	-0.21	0
15	15		0.112	0.050	0.025 E	.02	.04	-0.24	0
16	16		0.106	0.053	0.048 E	.01	.03	-0.24	0
17	17		-0.042	0.048	0.380 E	.02	.04	-0.23	0
18	18		-0.189	0.051	0.000 E	.02	.02	-0.22	0
19	19		0.077	0.051	0.126 E	.01	.04	-0.24	0
20	20		-0.172	0.050	0.001 E	.02	.03	-0.22	0

Program execution is completed.

Your output is stored on the file: a:\out1

ภาคผนวก ซ

ตัวอย่างผลการวิเคราะห์ DIF ด้วยวิธี GMH

```

A=[455 517 428 210 59]
B=A**A
C=[455 0 0 0 0;0 517 0 0 0;0 0 428 0 0;0 0 0 210 0;0 0 0 0 59]
D=C*1652
E=D-B
F=E/(1652*1652*1651)
G=F*826*826
H=inv(G)
I=[1.5 -26.5 -8 22 8.5]
J=I*H*I'

```

To get started, select "MATLAB Help" from the Help menu.

```

>> A=[455 517 428 210 59]
B=A**A
C=[455 0 0 0 0;0 517 0 0 0;0 0 428 0 0;0 0 0 210 0;0 0 0 0 59]
D=C*1652
E=D-B
F=E/(1652*1652*1651)
G=F*826*826
H=inv(G)
I=[1.5 -26.5 -8 22 8.5]
J=I*H*I'

```

A =

```

455 517 428 210 59

```

B =

```

207025    235235    194740    95550    26845
235235    267289    221276    108570    30503
194740    221276    183184    89880    25252
95550     108570    89880    44100    12390
26845     30503    25252    12390    3481

```

C =

```

455  0  0  0  0
0 517  0  0  0

```

0 0 428 0 0
0 0 0 210 0
0 0 0 0 59

D =

751660 0 0 0 0
0 854084 0 0 0
0 0 707056 0 0
0 0 0 346920 0
0 0 0 0 97468

E =

544635 -235235 -194740 -95550 -26845
-235235 586795 -221276 -108570 -30503
-194740 -221276 523872 -89880 -25252
-95550 -108570 -89880 302820 -12390
-26845 -30503 -25252 -12390 93987

F =

1.0e-003 *

0.1209 -0.0522 -0.0432 -0.0212 -0.0050
-0.0522 0.1302 -0.0491 -0.0241 -0.0068
-0.0432 -0.0491 0.1163 -0.0199 -0.0056
-0.0212 -0.0241 -0.0199 0.0672 -0.0027
-0.0060 -0.0068 -0.0056 -0.0027 0.0209

G =

82.4705 -35.6201 -29.4882 -14.4685 -4.0650
-35.6201 88.8545 -33.5064 -16.4400 -4.6189
-29.4882 -33.5064 79.3265 -13.6099 -3.8237
-14.4685 -16.4400 -13.6099 45.8540 -1.8761
-4.0650 -4.6189 -3.8237 -1.8761 14.2318

H =

-0.2264	-0.2352	-0.2352	-0.2352	-0.2352
-0.2352	-0.2274	-0.2352	-0.2352	-0.2352
-0.2352	-0.2352	-0.2258	-0.2352	-0.2352
-0.2352	-0.2352	-0.2352	-0.2161	-0.2352
-0.2352	-0.2352	-0.2352	-0.2352	-0.1674

I =

1.5000	-26.5000	-8.0000	22.0000	8.5000
--------	----------	---------	---------	--------

J =

18.6866

>

ภาคผนวก ฅ

ตัวอย่างคำสั่งการวิเคราะห์ด้วยโปรแกรมสำเร็จรูป PARSCALE
และผลการวิเคราะห์ DIF ด้วยวิธี Generalized Partial Credit Model

UTHAIWAN ANALYZE FILE ma20.DAT :using PARSCALE GPCM/FOR 20 ITEMS
>COMMENT

The data are 6 categories(0,1,2,3,4,5) in 20 Circle Items
for Mathematics Test/1000 Male Samples

```
>FILE    DFNAME='A.ma20.DAT', SAVE;
>SAVE    INF='ma20.INF', PAR='ma20.PAR', SCO='ma20.SCO';
>INPUT   NTOT=20, LENGTH=(20), NTEST=1, NFMT=1, MAXCAT=6;
(20A1)
>TEST    TNAME=CIRCLE1, NBLOCK=1, ITEMS=(1(1)20);
>BLOCK   NITEMS=20, NCAT=6, ORIGINAL=(0,1,2,3,4,5), MODIFIED=(1,2,3,4,5,6);
>CALIB   PARTIAL, SCALE=1.7, LOGISTIC, CYCLES=(50,1,1,1,1),
          NQPT=17, QRANGE=(-3.0,3.0), CRIT=0.001, SPRIOR, TPRIOR;
>SCORE   MLE, SMEAN=0.0, SSD=1.0, NAME=ma20;
```

1

PARSCALE RELEASE 22

MAXIMUM LIKELIHOOD ITEM ANALYSIS AND TEST SCORING: POLYTOMOUS MODEL

DISTRIBUTED BY

SCIENTIFIC SOFTWARE INTERNATIONAL, INC.

1525 East 53rd Street, Suite 530
 Chicago, IL 60615
 (800) 247-6113, (312) 684-4920

PROGRAM COPYRIGHT HELD BY SCIENTIFIC SOFTWARE INTERNATIONAL, INC. 1994

DISTRIBUTION OR USE UNAUTHORIZED BY SCIENTIFIC SOFTWARE INTERNATIONAL, INC. IS
 PROHIBITED.

1 *** POLYTOMOUS MODEL ITEM ANALYSER ***

*** PHASE 0 ***

UTHAIWAN ANALYZE FILE fe20.DAT :using PARSCALE GPCM FOR 20 ITEMS
 >COMMENT

--> FIND WARNING: 2 RECORDS NOT CONTAINING > IN COLUMN 1 HAVE BEEN SKIPPED

>FILE DFNAME='A:fe20.DAT',SAVE;

>SAVE INF='fe20.INF',PAR='fe20.PAR',SCO='fe20.SCO';

>INPUT NTOT=20,LENGTH=(20),NTEST=1,NFMT=1,MAXCAT=6;

SINGLE MAIN TEST IS USED.
 NUMBER OF ITEMS: 20

FORMAT OF DATA INPUT IS
(20A1)

>TEST TNAME=CIRCLE1,NBlock=1, ITEMS=(1(1)20);

BLOCK CARD: 1

>BLOCK NITEMS=20,NCAT=6,ORIGinaL=(0,1,2,3,4,5),MODified=(1,2,3,4,5,6);

>CALIB PARTIAL,SCALE=1.7,LOGISTIC,CYCLES=(50,1,1,1,1),

NQPT=17,QRANGE=(-3.0,3.0),CRIT=0.001,SPRIOR,TPRIOR;

MODEL SPECIFICATIONS

LOGISTIC -PARTIAL CREDIT ITEM RESPONSE MODEL IS SPECIFIED.
SCALE CONSTANT 1.70 FOR SLOPE PARAMETERS.

CALIBRATION PARAMETERS

MAXIMUM NUMBER OF EM CYCLES: 50

MAXIMUM INNER EM CYCLES: 1

MAXIMUM CATEGORY ESTIMATION CYCLES: 1

MAXIMUM ITEM PARAMETER ESTIMATION CYCLES: 1

MINIMUM CATEGORY AND ITEM ESTIMATION CYCLES 1

CONVERGENCE CRITERION FOR EM CYCLES: 0.0010

CONVERGENCE CRITERION FOR SLOPE: 0.0010

CONVERGENCE CRITERION FOR THRESHOLD: 0.0010

CONVERGENCE CRITERION FOR CATEGORY: 0.0010

CONVERGENCE CRITERION FOR GEUSSING 0.0010

ORDER OF INNER EM CYCLES: CATEGORY -ITEM PARAMETERS

ESTIMATION ACCELERATOR: YES

PRIOR DISTRIBUTION ON SLOPES: YES

PRIOR DISTRIBUTION ON THRESHOLD: YES

QUADRATURE POINTS AND PRIOR WEIGHTS (PROGRAM-GENERATED NORMAL APPROXIMATION)

	1	2	3	4	5
POINT	0.3000E+01	0.2625E+01	0.2250E+01	0.1875E+01	0.1500E+01
WEIGHT	0.1664E-02	0.4778E-02	0.1192E-01	0.2583E-01	0.4863E-01

	6	7	8	9	10
POINT	0.1125E+01	0.7500E+00	0.3750E+00	0.0000E+00	-0.3750E+00
WEIGHT	0.7956E-01	0.1131E+00	0.1396E+00	0.1498E+00	0.1396E+00

	11	12	13	14	15
POINT	-0.7500E+00	-0.1125E+01	-0.1500E+01	-0.1875E+01	-0.2250E+01
WEIGHT	0.1131E+00	0.7956E-01	0.4863E-01	0.2583E-01	0.1192E-01

	16	17
POINT	-0.2625E+01	-0.3000E+01
WEIGHT	0.4778E-02	0.1664E-02

TOTAL WEIGHT: 1.00000

MEAN : 0.00000

S.D. : 0.99238

PRIOR DISTRIBUTIONS ON ITEM PARAMETERS
(THRESHOLDS, NORMAL; SLOPES, LOG-NORMAL)

ITEM	THRESHOLDS				SLOPES		GUESSING
	MU	SIGMA	MU	SIGMA	MU	SIGMA	
0001	0.000	2.000	0.000	0.500	-	-	
0002	0.000	2.000	0.000	0.500	-	-	
0003	0.000	2.000	0.000	0.500	-	-	
0004	0.000	2.000	0.000	0.500	-	-	
0005	0.000	2.000	0.000	0.500	-	-	
0006	0.000	2.000	0.000	0.500	-	-	
0007	0.000	2.000	0.000	0.500	-	-	
0008	0.000	2.000	0.000	0.500	-	-	
0009	0.000	2.000	0.000	0.500	-	-	
0010	0.000	2.000	0.000	0.500	-	-	
0011	0.000	2.000	0.000	0.500	-	-	
0012	0.000	2.000	0.000	0.500	-	-	
0013	0.000	2.000	0.000	0.500	-	-	
0014	0.000	2.000	0.000	0.500	-	-	
0015	0.000	2.000	0.000	0.500	-	-	
0016	0.000	2.000	0.000	0.500	-	-	
0017	0.000	2.000	0.000	0.500	-	-	
0018	0.000	2.000	0.000	0.500	-	-	
0019	0.000	2.000	0.000	0.500	-	-	
0020	0.000	2.000	0.000	0.500	-	-	

>SCORE MLE, SMEAN=0.0, SSD=1.0, NAME=fe20;

PARAMETERS FOR SCORING AND TEST AND ITEM INFORMATION

METHOD OF SCORING SUBJECTS: MAXIMUM LIKELIHOOD
 TYPE OF PRIOR: NORMAL APPROXIMATION
 SCORES WRITTEN TO FILE fe20.SCO

SET NUMBER : 1
 SCORE NAME : fe20
 NUMBER OF ITEMS : 20
 RESCALE CONSTANT: MEAN = 0.00 S.D. = 1.00
 ITEMS : 1 2 3 4 5 6 7 8 9 10
 11 12 13 14 15 16 17 18 19 20

0001 0002 0003 0004 0005 0006 0007 0008 0009 0010
 0011 0012 0013 0014 0015 0016 0017 0018 0019 0020

FILE ASSIGNMENT AND DISPOSITION

[INPUT FILES]

SUBJECT DATA INPUT FILE A:fe20.DAT
 SINGLE-SUBJECT DATA
 NO CASE WEIGHTS

[OUTPUT FILES]

ITEM PARAMETERS FILE fe20.PAR
 SUBJECT SCALE-SCORE FILE fe20.SCO
 ITEM-TEST INFORMATION FILE fe20.INF

[SCRATCH FILES]

PARSCALE SYSTEM BINARY DATA FILE fe201.MFL
 TEMPORARY FILE fe201.T99
 TEMPORARY FILE fe201.T98

INPUT AND RECODED RESPONSE OF FIRST AND SECOND OBSERVATIONS

OBSERVATION # 1
 ID: 334212303332243244134314233414
 RESPONSE: 14332235232341234431
 BLOCK NAME: BLOCK 01
 RECODED RESPONSE: 25443346343452345542

OBSERVATION # 2
 ID: 332124231303344544242214235232

RESPONSE: 12342034455453233423
BLOCK NAME: BLOCK 01
RECODED RESPONSE: 234531145566564344534

MAINTEST: CIRCLE1

333 OBSERVATIONS READ FROM FILE: A:fe20.DAT
333 OBSERVATIONS WRITTEN TO FILE: fe201.MFL

3772 BYTES OF NUMERICAL WORKSPACE USED OF 33537920 AVAILABLE IN PHASE-0
1252 BYTES OF CHARACTER WORKSPACE USED OF 80000 AVAILABLE IN PHASE-0

[PHASE 0]

CURRENT DATE: 03/14/03
CURRENT TIME: 12:50:33.58

PARSCALE V2.2

MAXIMUM LIKELIHOOD ITEM ANALYSIS AND TEST SCORING: POLYTOMOUS MODEL

1 *** POLYTOMOUS MODEL ITEM ANALYSER ***

... PHASE 1 ...

```
UTHAIWAN ANALYZE FILE fe20.DAT :using PARSCALE GPCM/FOR 20 ITEMS
>COMMENT
```

MAINTEST: CIRCLE:

333 OBS(WEIGHTS: 333.000) WERE READ FROM fe201MFL

BLOCK NO.: 1 NAME: BLOCK 01

ITEM NAME	TOTAL	NOT-PRESENT	OMIT	1	2	3
	4	5	6			
0001						
FREQ.	333.00	0.00	0.00	6.00	61.00	114.00
PERCENT				0.00	0.00	1.80
				18.32	34.23	
CUM FREQ.						
CUM FREQ.						
←						
0002						
FREQ.	333.00	0.00	0.00	3.00	18.00	54.00
PERCENT				0.00	0.00	0.90
				5.41	16.22	
CUM FREQ.						

FREQ.	333.00	0.00	0.00	16.00	36.00	93.00			
							106.00	61.00	21.00
PERCENT				0.00	0.00	4.80	10.81	27.93	
							31.83	18.32	6.31
CUM FREQ.							16.00	52.00	145.00
							251.00	312.00	333.00
CUM FREQ.							333.00	317.00	281.00
							183.00	82.00	21.00
0012									
FREQ.	333.00	0.00	0.00	7.00	34.00	72.00			
							102.00	57.00	61.00
PERCENT				0.00	0.00	2.10	10.21	21.62	
							30.63	17.12	18.32
CUM FREQ.							7.00	41.00	113.00
							215.00	272.00	333.00
CUM FREQ.							333.00	326.00	292.00
							220.00	118.00	61.00
0013									
FREQ.	333.00	0.00	0.00	5.00	31.00	93.00			
							87.00	59.00	58.00
PERCENT				0.00	0.00	1.50	9.31	27.93	
							26.13	17.72	17.42
CUM FREQ.							5.00	36.00	129.00
							216.00	275.00	333.00
CUM FREQ.							333.00	328.00	297.00
							204.00	117.00	58.00
0014									
FREQ.	333.00	0.00	0.00	8.00	55.00	107.00			
							91.00	45.00	27.00
PERCENT				0.00	0.00	2.40	16.52	32.13	
							27.33	13.51	8.11
CUM FREQ.							8.00	63.00	170.00
							261.00	306.00	333.00
CUM FREQ.							333.00	325.00	270.00
							163.00	72.00	27.00
0015									
FREQ.	333.00	0.00	0.00	5.00	40.00	130.00			
							95.00	41.00	22.00
PERCENT				0.00	0.00	1.50	12.01	39.04	

8 0008		4.760		79.544		0.369		0.432
		(1.127)		(10.319)		0.397		
9 0009		3.520		79.544		0.277		0.299
		(1.184)		(10.319)		0.287		
10 0010		4.036		79.544		0.297		0.326
		(1.062)		(10.319)		0.310		
11 0011		3.670		79.544		0.458		0.537
		(1.213)		(10.319)		0.473		
12 0012		4.054		79.544		0.534		0.669
		(1.310)		(10.319)		0.556		
13 0013		4.015		79.544		0.526		0.656
		(1.289)		(10.319)		0.549		
14 0014		3.574		79.544		0.494		0.597
		(1.215)		(10.319)		0.513		
15 0015		3.580		79.544		0.425		0.496
		(1.103)		(10.319)		0.445		
16 0016		3.892		79.544		0.401		0.457
		(1.181)		(10.319)		0.416		
17 0017		4.282		79.544		0.408		0.471
		(1.062)		(10.319)		0.426		
18 0018		4.330		79.544		0.531		0.668
		(1.203)		(10.319)		0.555		
19 0019		3.610		79.544		0.412		0.473
		(1.261)		(10.319)		0.427		
20 0020		3.892		79.544		0.471		0.562
		(1.126)		(10.319)		0.490		

CATEGORY		SCORING		MEAN		S.D.		PARAMETER
1		1.000		73.153		6.768		0.000
2		2.000		74.863		7.441		2.504
3		3.000		76.392		7.671		1.428
4		4.000		77.761		8.153		0.491
5		5.000		81.676		9.731		-0.398
6		6.000		90.554		13.372		-1.148

5444 BYTES OF NUMERICAL WORKSPACE USED OF 33537920 AVAILABLE IN PHASE-1
280 BYTES OF CHARACTER WORKSPACE USED OF 80000 AVAILABLE IN PHASE-1

[PHASE 1]

CURRENT DATE : 03/14/03
STARTING TIME: 12:50:37
CURRENT TIME : 12:50:37

PARSCALE V22

MAXIMUM LIKELIHOOD ITEM ANALYSIS AND TEST SCORING: POLYTOMOUS MODEL

1 *** POLYTOMOUS MODEL ITEM ANALYSER ***

*** PHASE 2 ***

UTHAIWAN ANALYZE FILE fe2.DAT :using PARSCALE GPCM/ FOR 20 ITEMS
>COMMENT

MAINTEST: CIRCLE1

CALIBRATION OF MAINTEST
CIRCLE1

[E-M CYCLES] PARTIAL CREDIT MODEL

STEP AND ITEM PARAMETERS AFTER CYCLE 0

ITEM TYPE 1 BLOCK 01

SCORING FUNCTION : 1.000 2.000 3.000 4.000 5.000 6.000

STEP PARAMETER : 0.000 2.504 1.428 0.491 -0.398 -1.148

S.E. : 0.000 0.000 0.000 0.000 0.000 0.000

0001 ITEM-CATEGORY: 0.000 -2.504 -1.428 -0.491 0.398 1.148

INTERCEPT : 0.000 1.571 0.896 0.308 -0.250 -0.720

0002 ITEM-CATEGORY: 0.000 -2.504 -1.428 -0.491 0.398 1.148

```

INTERCEPT      : 0.000 1.547 0.882 0.304 -0.246 -0.709

0003 ITEM-CATEGORY: 0.000 -2.504 -1.428 -0.491 0.398 1.148

INTERCEPT      : 0.000 1.001 0.571 0.196 -0.159 -0.459

0004 ITEM-CATEGORY: 0.000 -2.504 -1.428 -0.491 0.398 1.148

INTERCEPT      : 0.000 1.064 0.606 0.209 -0.169 -0.488

0005 ITEM-CATEGORY: 0.000 -2.504 -1.428 -0.491 0.398 1.148

INTERCEPT      : 0.000 1.584 0.903 0.311 -0.252 -0.726

0006 ITEM-CATEGORY: 0.000 -2.504 -1.428 -0.491 0.398 1.148

INTERCEPT      : 0.000 1.258 0.717 0.247 -0.200 -0.577

0007 ITEM-CATEGORY: 0.000 -2.504 -1.428 -0.491 0.398 1.148

INTERCEPT      : 0.000 1.448 0.825 0.284 -0.230 -0.664

0008 ITEM-CATEGORY: 0.000 -2.504 -1.428 -0.491 0.398 1.148

INTERCEPT      : 0.000 1.082 0.617 0.212 -0.172 -0.496

0009 ITEM-CATEGORY: 0.000 -2.504 -1.428 -0.491 0.398 1.148

INTERCEPT      : 0.000 0.749 0.427 0.147 -0.119 -0.343

0010 ITEM-CATEGORY: 0.000 -2.504 -1.428 -0.491 0.398 1.148

INTERCEPT      : 0.000 0.816 0.465 0.160 -0.130 -0.374

0011 ITEM-CATEGORY: 0.000 -2.504 -1.428 -0.491 0.398 1.148

INTERCEPT      : 0.000 1.345 0.767 0.264 -0.214 -0.617

0012 ITEM-CATEGORY: 0.000 -2.504 -1.428 -0.491 0.398 1.148

INTERCEPT      : 0.000 1.675 0.955 0.329 -0.266 -0.768

0013 ITEM-CATEGORY: 0.000 -2.504 -1.428 -0.491 0.398 1.148

INTERCEPT      : 0.000 1.644 0.937 0.323 -0.261 -0.753

0014 ITEM-CATEGORY: 0.000 -2.504 -1.428 -0.491 0.398 1.148

INTERCEPT      : 0.000 1.496 0.853 0.294 -0.238 -0.686

0015 ITEM-CATEGORY: 0.000 -2.504 -1.428 -0.491 0.398 1.148

INTERCEPT      : 0.000 1.243 0.709 0.244 -0.198 -0.570

0016 ITEM-CATEGORY: 0.000 -2.504 -1.428 -0.491 0.398 1.148

INTERCEPT      : 0.000 1.145 0.653 0.225 -0.182 -0.525

0017 ITEM-CATEGORY: 0.000 -2.504 -1.428 -0.491 0.398 1.148

INTERCEPT      : 0.000 1.181 0.673 0.232 -0.188 -0.541

0018 ITEM-CATEGORY: 0.000 -2.504 -1.428 -0.491 0.398 1.148

INTERCEPT      : 0.000 1.672 0.953 0.328 -0.266 -0.766

```

0019 ITEM-CATEGORY: 0.000 -2.504 -1.428 -0.491 0.398 1.148

INTERCEPT : 0.000 1.183 0.675 0.232 -0.188 -0.543

0020 ITEM-CATEGORY: 0.000 -2.504 -1.428 -0.491 0.398 1.148

INTERCEPT : 0.000 1.407 0.802 0.276 -0.224 -0.645

ITEM	SLOPE	S.E.	THRESHOLD	S.E.	GUESSING	S.E.
0001	0.627	0.000	0.000	0.000	0.000	0.000
0002	0.618	0.000	0.000	0.000	0.000	0.000
0003	0.400	0.000	0.000	0.000	0.000	0.000
0004	0.425	0.000	0.000	0.000	0.000	0.000
0005	0.633	0.000	0.000	0.000	0.000	0.000
0006	0.502	0.000	0.000	0.000	0.000	0.000
0007	0.578	0.000	0.000	0.000	0.000	0.000
0008	0.432	0.000	0.000	0.000	0.000	0.000
0009	0.299	0.000	0.000	0.000	0.000	0.000
0010	0.326	0.000	0.000	0.000	0.000	0.000
0011	0.537	0.000	0.000	0.000	0.000	0.000
0012	0.669	0.000	0.000	0.000	0.000	0.000
0013	0.656	0.000	0.000	0.000	0.000	0.000
0014	0.597	0.000	0.000	0.000	0.000	0.000
0015	0.496	0.000	0.000	0.000	0.000	0.000
0016	0.457	0.000	0.000	0.000	0.000	0.000
0017	0.471	0.000	0.000	0.000	0.000	0.000
0018	0.668	0.000	0.000	0.000	0.000	0.000
0019	0.473	0.000	0.000	0.000	0.000	0.000
0020	0.562	0.000	0.000	0.000	0.000	0.000

LARGEST CHANGE 0.00000

PARAMETER	MEAN	STN	DEV	N
SLOPE	0.521	0.111	20	
LOG(SLOPE)	-0.675	0.232	20	
THRESHOLD	0.000	0.000	20	
GUESSING	0.000	0.000	0	

-2 LOG LIKELIHOOD = 21082.996

STEP AND ITEM PARAMETERS AFTER CYCLE 1

LARGEST CHANGE 1.45246

-2 LOG LIKELIHOOD = 20744.302

STEP AND ITEM PARAMETERS AFTER CYCLE 2

LARGEST CHANGE 3.15036

-2 LOG LIKELIHOOD = 20225.104

STEP AND ITEM PARAMETERS AFTER CYCLE 3

LARGEST CHANGE 0.54338

-2 LOG LIKELIHOOD = 20306.670

STEP AND ITEM PARAMETERS AFTER CYCLE 4

LARGEST CHANGE 0.34629

-2 LOG LIKELIHOOD = 20103.417

STEP AND ITEM PARAMETERS AFTER CYCLE 5

LARGEST CHANGE 1.02828

-2 LOG LIKELIHOOD = 20008.653

STEP AND ITEM PARAMETERS AFTER CYCLE 6

LARGEST CHANGE 0.26652

-2 LOG LIKELIHOOD = 19991.672

STEP AND ITEM PARAMETERS AFTER CYCLE 7

LARGEST CHANGE 0.10420

-2 LOG LIKELIHOOD = 19982.525

STEP AND ITEM PARAMETERS AFTER CYCLE 8

LARGEST CHANGE 0.05965

-2 LOG LIKELIHOOD = 19981.100

STEP AND ITEM PARAMETERS AFTER CYCLE 9

LARGEST CHANGE 0.15837

-2 LOG LIKELIHOOD = 19980.630

STEP AND ITEM PARAMETERS AFTER CYCLE 10

LARGEST CHANGE 0.10011

-2 LOG LIKELIHOOD = 19977.963

STEP AND ITEM PARAMETERS AFTER CYCLE 11

LARGEST CHANGE 0.02318

-2 LOG LIKELIHOOD = 19977.450

STEP AND ITEM PARAMETERS AFTER CYCLE 12

LARGEST CHANGE 0.01888

-2 LOG LIKELIHOOD = 19976.556

STEP AND ITEM PARAMETERS AFTER CYCLE 13

LARGEST CHANGE 0.02161

-2 LOG LIKELIHOOD = 19975.881

STEP AND ITEM PARAMETERS AFTER CYCLE 14

LARGEST CHANGE 0.00955

-2 LOG LIKELIHOOD = 19975.575

STEP AND ITEM PARAMETERS AFTER CYCLE 15

LARGEST CHANGE 0.01151

-2 LOG LIKELIHOOD = 19975.172

STEP AND ITEM PARAMETERS AFTER CYCLE 16

LARGEST CHANGE 0.01090

-2 LOG LIKELIHOOD = 19974.920

STEP AND ITEM PARAMETERS AFTER CYCLE 17

LARGEST CHANGE 0.00533

-2 LOG LIKELIHOOD = 19974.745

STEP AND ITEM PARAMETERS AFTER CYCLE 18

LARGEST CHANGE 0.01106

-2 LOG LIKELIHOOD = 19974.562

STEP AND ITEM PARAMETERS AFTER CYCLE 19

LARGEST CHANGE 0.00665

-2 LOG LIKELIHOOD = 19974.412

STEP AND ITEM PARAMETERS AFTER CYCLE 20

LARGEST CHANGE 0.00247

-2 LOG LIKELIHOOD = 19974.273

STEP AND ITEM PARAMETERS AFTER CYCLE 21

LARGEST CHANGE 0.00376

-2 LOG LIKELIHOOD = 19974.152

STEP AND ITEM PARAMETERS AFTER CYCLE 22

LARGEST CHANGE 0.00281

-2 LOG LIKELIHOOD = 19974.083

STEP AND ITEM PARAMETERS AFTER CYCLE 23

LARGEST CHANGE 0.00151

-2 LOG LIKELIHOOD = 19973.999

STEP AND ITEM PARAMETERS AFTER CYCLE 24

LARGEST CHANGE 0.00263

-2 LOG LIKELIHOOD = 19973.938

STEP AND ITEM PARAMETERS AFTER CYCLE 25

LARGEST CHANGE 0.00226

-2 LOG LIKELIHOOD = 19973.878

STEP AND ITEM PARAMETERS AFTER CYCLE 26

LARGEST CHANGE 0.00133

-2 LOG LIKELIHOOD = 19973.823

STEP AND ITEM PARAMETERS AFTER CYCLE 27

LARGEST CHANGE 0.00315

-2 LOG LIKELIHOOD = 19973.787

STEP AND ITEM PARAMETERS AFTER CYCLE 28

LARGEST CHANGE 0.00117

-2 LOG LIKELIHOOD = 19973.743

STEP AND ITEM PARAMETERS AFTER CYCLE 29

ITEM TYPE 1 BLOCK 01

```

SCORING FUNCTION : 1.000 2.000 3.000 4.000 5.000 6.000

STEP PARAMTER : 0.000 3.103 1.292 -0.262 -1.619 -2.514

S.E. : 0.000 0.183 0.082 0.059 0.062 0.081

0001 ITEM-CATEGORY: 0.000 -3.275 -1.463 0.090 1.448 2.342

INTERCEPT : 0.000 1.254 0.560 -0.035 -0.554 -0.897

0002 ITEM-CATEGORY: 0.000 -4.488 -2.676 -1.123 0.234 1.129

INTERCEPT : 0.000 1.792 1.069 0.449 -0.094 -0.451

0003 ITEM-CATEGORY: 0.000 -4.197 -2.386 -0.832 0.525 1.419

INTERCEPT : 0.000 1.522 0.865 0.302 -0.190 -0.515

0004 ITEM-CATEGORY: 0.000 -6.081 -4.270 -2.717 -1.359 -0.465

INTERCEPT : 0.000 1.313 0.922 0.587 0.294 0.100

0005 ITEM-CATEGORY: 0.000 -3.557 -1.745 -0.192 1.166 2.060

INTERCEPT : 0.000 1.094 0.537 0.059 -0.359 -0.634

0006 ITEM-CATEGORY: 0.000 -3.883 -2.071 -0.518 0.839 1.734

INTERCEPT : 0.000 1.140 0.608 0.152 -0.246 -0.509

0007 ITEM-CATEGORY: 0.000 -3.912 -2.101 -0.548 0.810 1.704

INTERCEPT : 0.000 1.598 0.858 0.224 -0.331 -0.696

0008 ITEM-CATEGORY: 0.000 -5.514 -3.703 -2.149 -0.792 0.102

INTERCEPT : 0.000 1.373 0.922 0.535 0.197 -0.025

0009 ITEM-CATEGORY: 0.000 -3.293 -1.482 0.071 1.429 2.323

INTERCEPT : 0.000 0.858 0.386 -0.019 -0.372 -0.605

0010 ITEM-CATEGORY: 0.000 -4.103 -2.292 -0.739 0.619 1.513

INTERCEPT : 0.000 1.362 0.761 0.245 -0.205 -0.502

0011 ITEM-CATEGORY: 0.000 -3.538 -1.726 -0.173 1.184 2.079

INTERCEPT : 0.000 1.094 0.534 0.053 -0.366 -0.643

0012 ITEM-CATEGORY: 0.000 -4.155 -2.343 -0.790 0.567 1.462

INTERCEPT : 0.000 1.208 0.682 0.230 -0.165 -0.425

0013 ITEM-CATEGORY: 0.000 -4.082 -2.270 -0.717 0.640 1.535

INTERCEPT : 0.000 1.257 0.699 0.221 -0.197 -0.473

0014 ITEM-CATEGORY: 0.000 -3.389 -1.578 -0.024 1.333 2.227

INTERCEPT : 0.000 1.156 0.538 0.008 -0.455 -0.760

0015 ITEM-CATEGORY: 0.000 -3.403 -1.592 -0.039 1.319 2.213

```

```
INTERCEPT      : 0.000 1.355 0.634 0.015 -0.525 -0.881

0016 ITEM-CATEGORY: 0.000 -3.885 -2.074 -0.521 0.837 1.731

INTERCEPT      : 0.000 1.215 0.648 0.163 -0.262 -0.541

0017 ITEM-CATEGORY: 0.000 -4.449 -2.638 -1.084 0.273 1.167

INTERCEPT      : 0.000 1.801 1.068 0.439 -0.111 -0.473

0018 ITEM-CATEGORY: 0.000 -4.552 -2.740 -1.187 0.170 1.065

INTERCEPT      : 0.000 1.608 0.968 0.419 -0.060 -0.376

0019 ITEM-CATEGORY: 0.000 -3.442 -1.631 -0.078 1.280 2.174

INTERCEPT      : 0.000 0.941 0.446 0.021 -0.350 -0.594

0020 ITEM-CATEGORY: 0.000 -3.870 -2.058 -0.505 0.852 1.747

INTERCEPT      : 0.000 1.517 0.807 0.198 -0.334 -0.685
```

ITEM	SLOPE		S.E. THRESHOLD		S.E. GUESSING		S.E.
0001	0.383	0.032	-0.172	0.084	0.000	0.000	
0002	0.399	0.037	-1.385	0.083	0.000	0.000	
0003	0.363	0.033	-1.094	0.087	0.000	0.000	
0004	0.216	0.031	-2.979	0.237	0.000	0.000	
0005	0.308	0.028	-0.454	0.095	0.000	0.000	
0006	0.294	0.027	-0.780	0.098	0.000	0.000	
0007	0.408	0.035	-0.809	0.080	0.000	0.000	
0008	0.249	0.031	-2.411	0.165	0.000	0.000	
0009	0.260	0.025	-0.191	0.106	0.000	0.000	
0010	0.332	0.031	-1.000	0.092	0.000	0.000	
0011	0.309	0.028	-0.435	0.095	0.000	0.000	
0012	0.291	0.028	-1.052	0.101	0.000	0.000	
0013	0.308	0.029	-0.979	0.096	0.000	0.000	
0014	0.341	0.029	-0.286	0.089	0.000	0.000	
0015	0.398	0.033	-0.301	0.082	0.000	0.000	
0016	0.313	0.029	-0.783	0.094	0.000	0.000	
0017	0.405	0.037	-1.346	0.082	0.000	0.000	
0018	0.353	0.034	-1.449	0.092	0.000	0.000	
0019	0.273	0.025	-0.339	0.102	0.000	0.000	
0020	0.392	0.034	-0.767	0.082	0.000	0.000	

LARGEST CHANGE 0.00064

PARAMETER	MEAN	STN	DEV	N
SLOPE	0.330	0.057	20	
LOG(SLOPE)	-1.124	0.180	20	
THRESHOLD	-0.950	0.723	20	
GUESSING	0.000	0.000	0	

QUADRATURE POINTS AND POSTERIOR WEIGHTS:

	1	2	3	4	5
POINT		0.3000E+01	0.2625E+01	0.2250E+01	0.1875E+01 0.1500E+01
WEIGHT		0.8588E+02	0.1044E+01	0.1078E+01	0.1276E+01 0.2168E+01

	6	7	8	9	10
POINT		0.1125E+01	0.7500E+00	0.3750E+00	0.0000E+00 -0.3750E+00
WEIGHT		0.4241E+01	0.7652E+01	0.1323E+00	0.2028E+00 0.2231E+00

	11	12	13	14	15
POINT		-0.7500E+00	-0.1125E+01	-0.1500E+01	-0.1875E+01 -0.2250E+01
WEIGHT		0.1638E+00	0.7362E+01	0.1848E+01	0.2440E+02 0.1597E+03

	16	17
POINT		-0.2625E+01 -0.3000E+01
WEIGHT		0.4904E+05 0.6869E+07

TOTAL WEIGHT: 1.00000
MEAN : 0.00000
S.D. : 0.99238

106360 BYTES OF NUMERICAL WORKSPACE USED OF 33537920 AVAILABLE IN PHASE-2
280 BYTES OF CHARACTER WORKSPACE USED OF 80000 AVAILABLE IN PHASE-2

[PHASE 2]

CURRENT DATE : 03/14/03
STARTING TIME : 12:50:40
CURRENT TIME : 12:50:46

PARSCALE V22

MAXIMUM LIKELIHOOD ITEM ANALYSIS AND TEST SCORING: POLYTOMOUS MODEL

1 *** POLYTOMOUS MODEL ITEM ANALYSER ***

*** PHASE 3 ***

UTHAIWAN ANALYZE FILE fe20.DAT using PARSCALE GPCM/ FOR 20 ITEMS
>COMMENT

MAINTEST: CIRCLE1

1

.....

SCORING OF MAIN TEST
CIRCLE1

.....

SCORING FUNCTION FOR SCALING

BLOCK: 1 BLOCK 01

- 1 1.000
- 2 2.000
- 3 3.000
- 4 4.000
- 5 5.000
- 6 6.000

MLE SUBJECT ESTIMATION

MEANS AND STANDARD DEVIATIONS OF ABILITY DISTRIBUTIONS BEFORE RESCALING

SCORE NAME	STANDARD DEVIATION	MEAN	TOTAL FREQUENCIES
fe20	0.936 -0.015	333.00	

RESCALING DONE WITH RESPECT TO USER SUPPLIED LINEAR TRANSFORMATION

SCORE NAME	SCALING CONSTANT	LOCATION CONSTANT	TOTAL FREQUENCIES
---------------	---------------------	----------------------	----------------------

fe20	1.069	0.016	333.00
------	-------	-------	--------

SUBJECT IDENTIFICATION		WEIGHT/FREQUENCY		
SCORE NAME	MEAN CATEGORY	ATTEMPTS	ABILITY	SE.

334212303332243 44	133	1.00		
1 fe20		3.75	1.00	-0.4317 0.3013

332124231303344 44	235	1.00		
1 fe20		4.10	1.00	0.0780 0.3015

132133343415231 24	333	1.00		
1 fe20		4.20	1.00	0.1613 0.3020

MEANS AND STANDARD DEVIATIONS OF ABILITY DISTRIBUTIONS AFTER RESCALING

SCORE NAME	STANDARD DEVIATION	MEAN	TOTAL FREQUENCIES
---------------	-----------------------	------	----------------------

fe20	1.000	0.000	333.00
------	-------	-------	--------

4568 BYTES OF NUMERICAL WORKSPACE USED OF 33537920 AVAILABLE IN PHASE-3
292 BYTES OF CHARACTER WORKSPACE USED OF 80000 AVAILABLE IN PHASE-3

[PHASE 3]

CURRENT DATE : 03/14/03
STARTING TIME: 12:50:49
CURRENT TIME : 12:51:01

ภาคผนวก ก

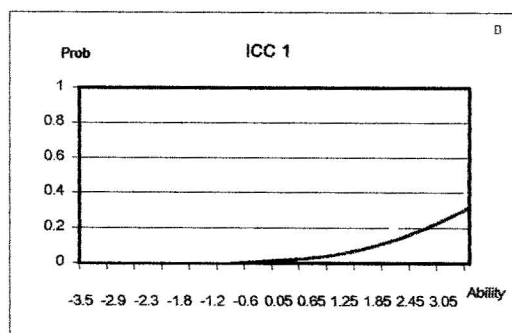
โค้งคุณลักษณะข้อสอบ

[Item Characteristic Curve : ICC]

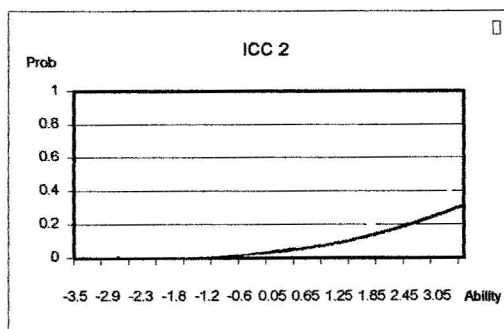
- แบบทดสอบขนาด 40 ข้อ
- แบบทดสอบขนาด 30 ข้อ
- แบบทดสอบขนาด 20 ข้อ

โค้งคุณลักษณะข้อสอบ (ICC) ที่วิเคราะห์ด้วยวิธี GPCM ในแบบทดสอบขนาด 40 ข้อ
เมื่อกลุ่มตัวอย่างเพศชาย ต่อ เพศหญิง เท่ากับ 1,000 คนต่อ 1,000 คน

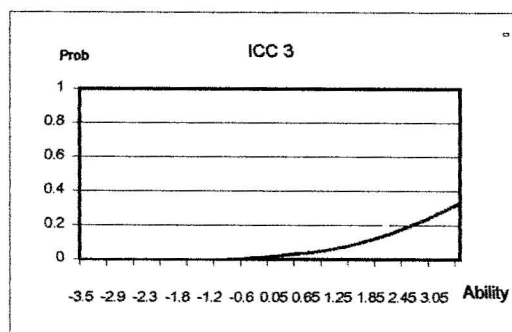
ข้อ 1



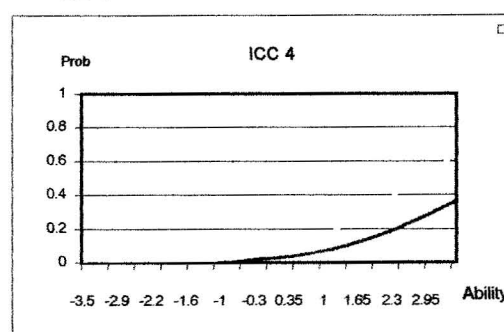
ข้อ 2



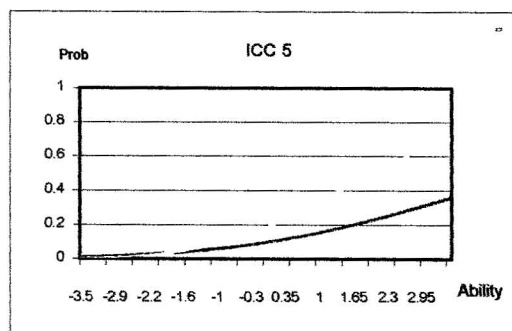
ข้อ 3



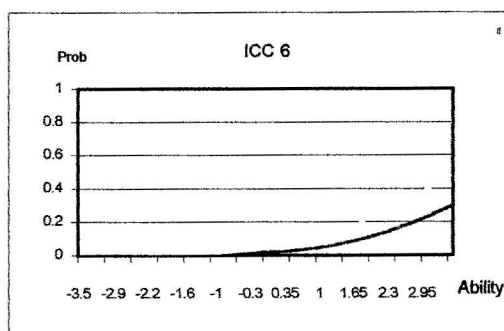
ข้อ 4



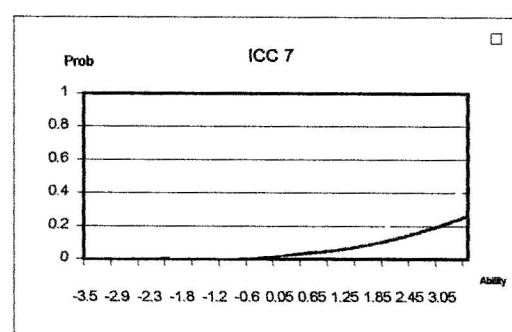
ข้อ 5



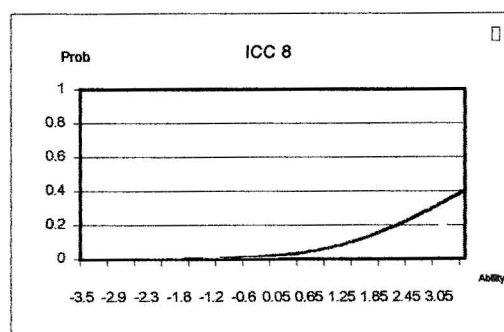
ข้อ 6



ข้อ 7



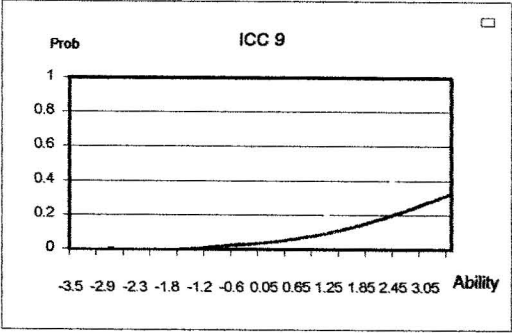
ข้อ 8



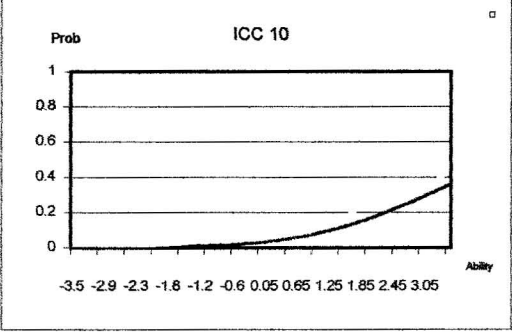
ชาย

หญิง

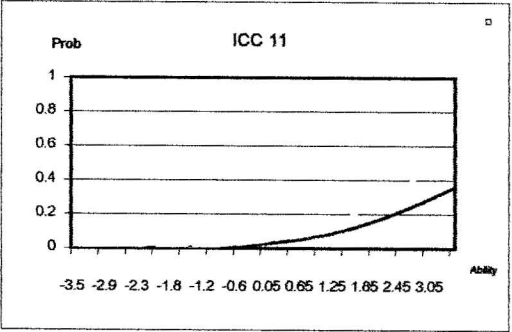
รูป 9



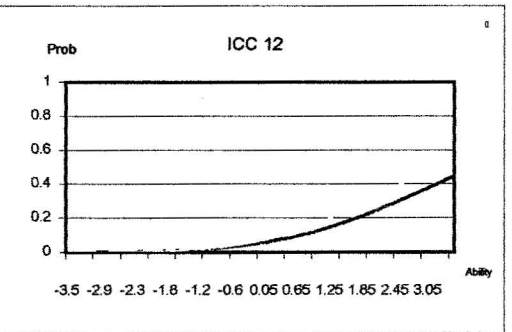
รูป 10



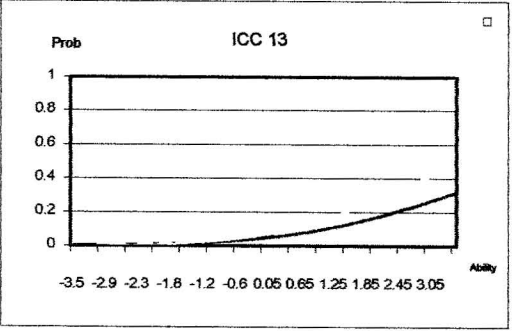
รูป 11



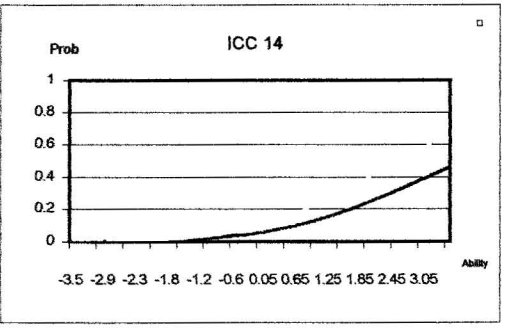
รูป 12



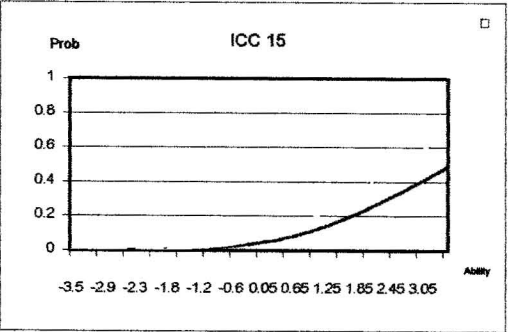
รูป 13



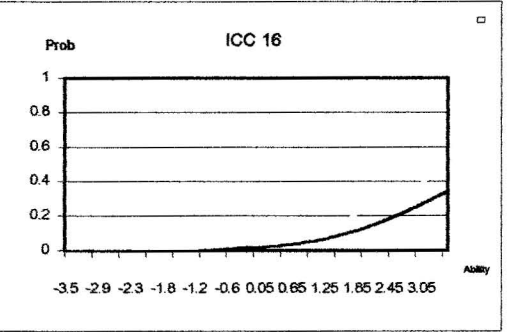
รูป 14



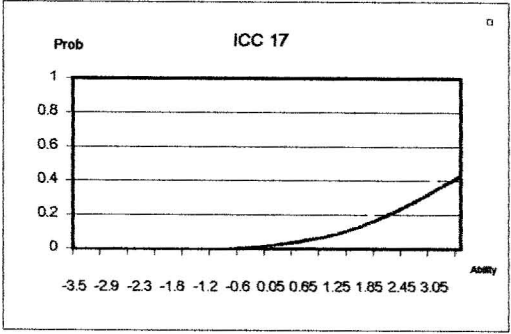
รูป 15



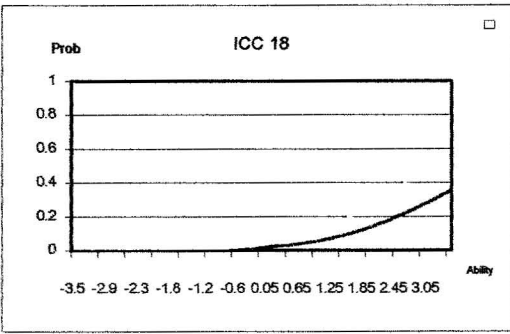
รูป 16



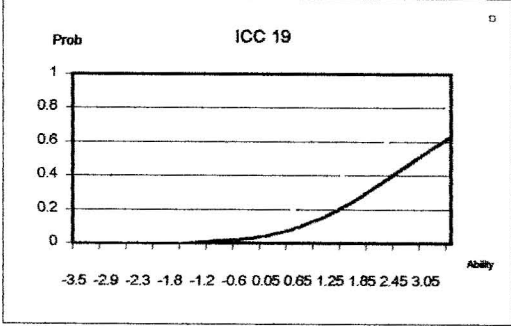
ข้อ 17



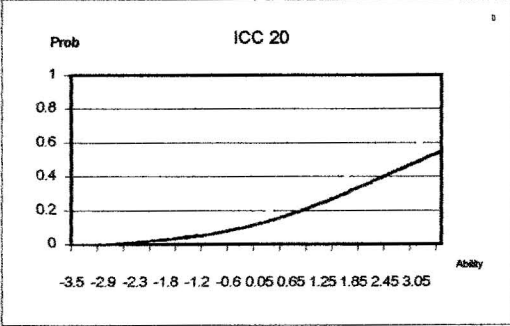
ข้อ 18



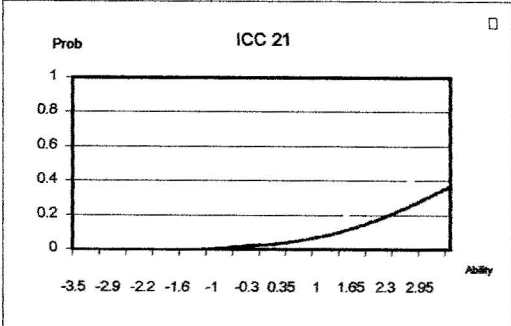
ข้อ 19



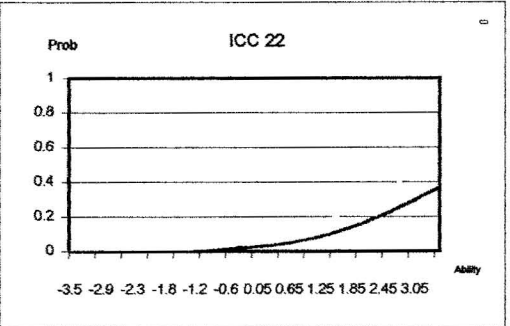
ข้อ 20



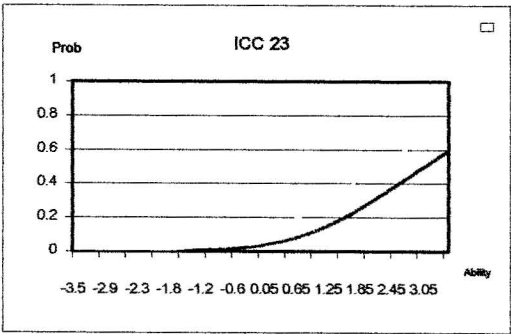
ข้อ 21



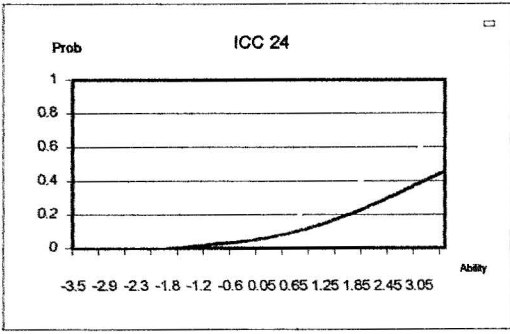
ข้อ 22



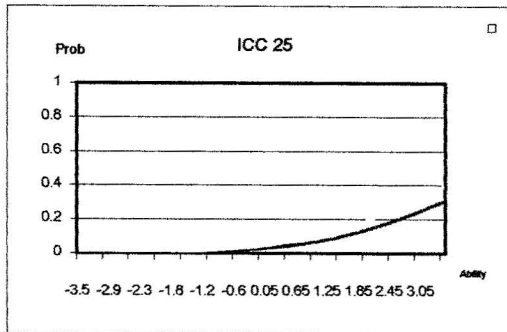
ข้อ 23



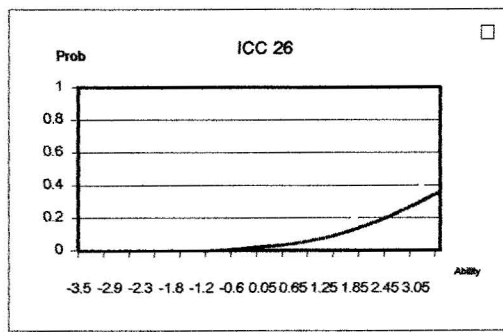
ข้อ 24



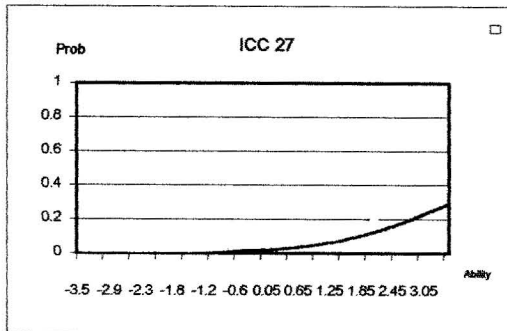
ข้อ 25



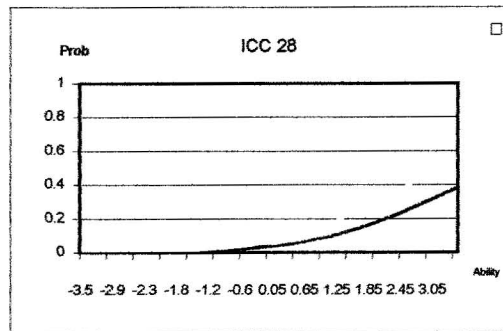
ข้อ 26



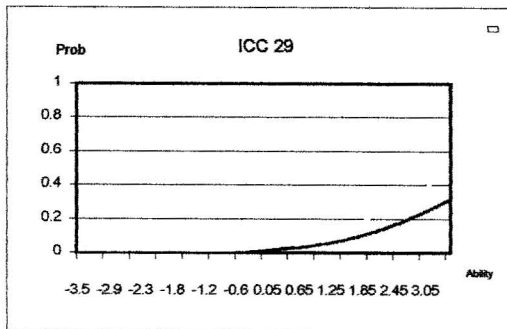
ข้อ 27



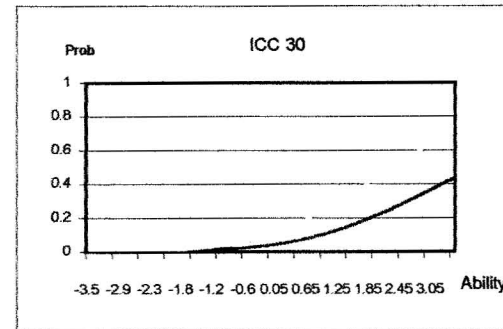
ข้อ 28



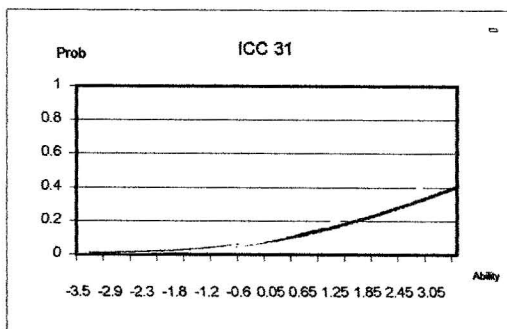
ข้อ 29



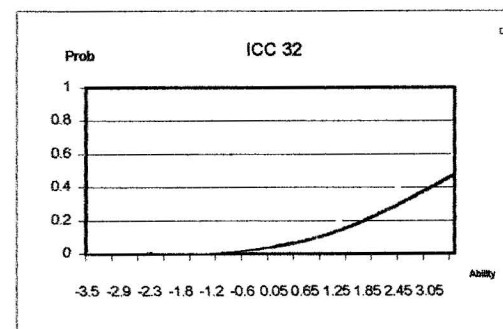
ข้อ 30



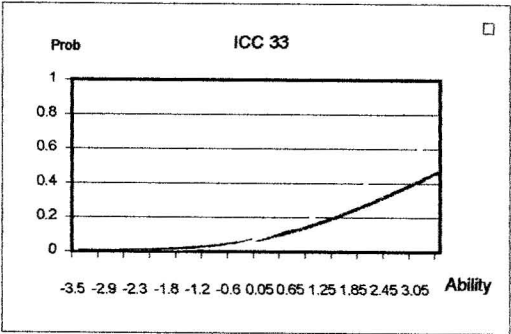
ข้อ 31



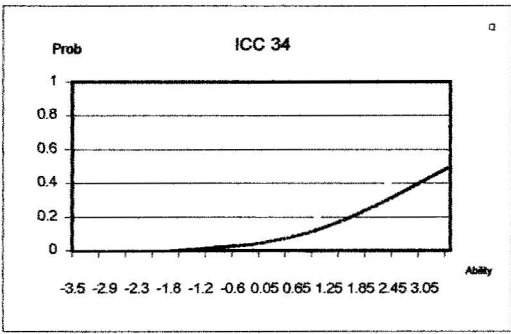
ข้อ 32



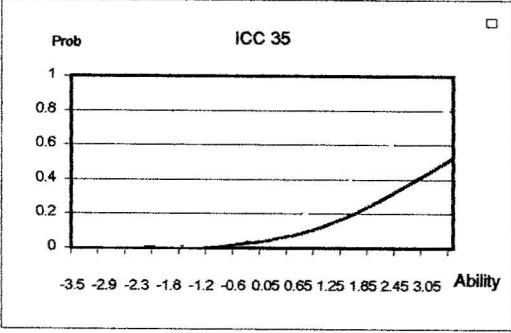
ข้อ 33



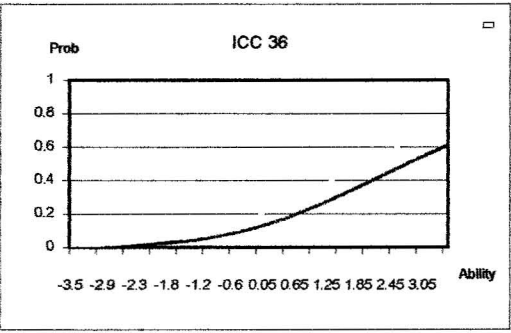
ข้อ 34



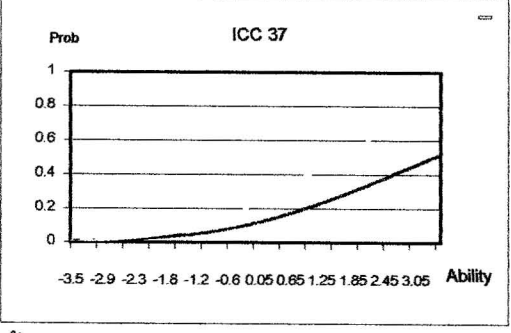
ข้อ 35



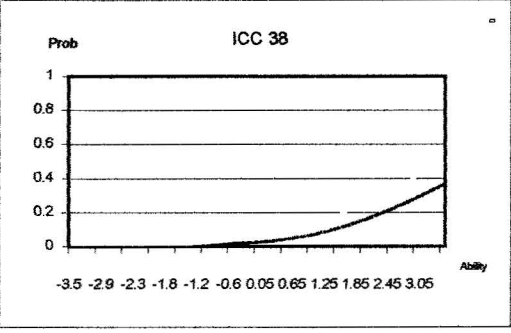
ข้อ 36



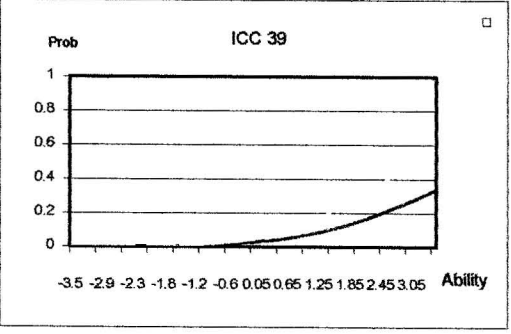
ข้อ 37



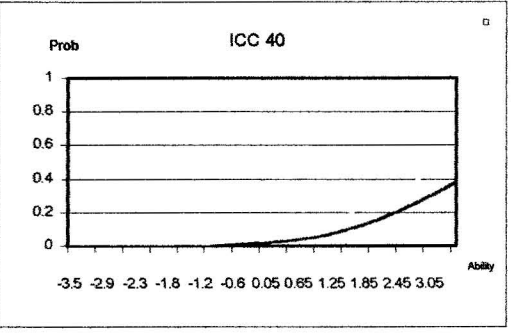
ข้อ 38



ข้อ 39



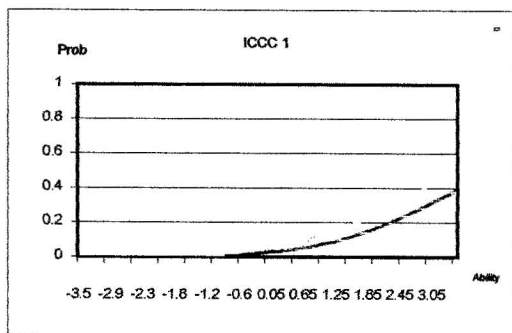
ข้อ 40



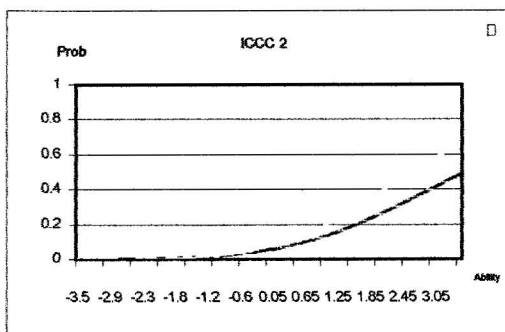
ได้คุณลักษณะข้อสอบ (ICC) ที่วิเคราะห์ด้วยวิธี GPCM ในแบบทดสอบขนาด 30 ข้อ
เมื่อกลุ่มตัวอย่างเพศชาย ต่อ เพศหญิง เท่ากับ 1,000 คน ต่อ 1,000 คน

186

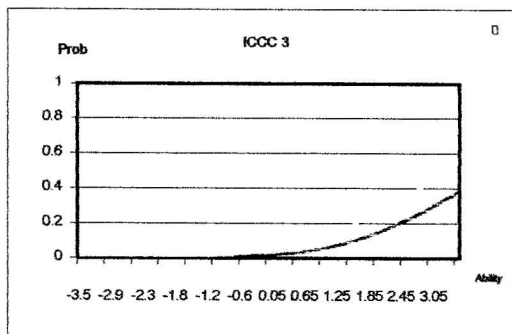
ข้อ 1



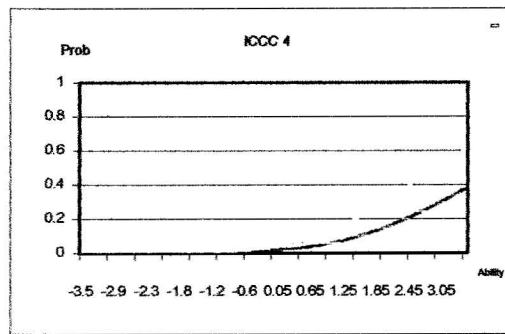
ข้อ 2



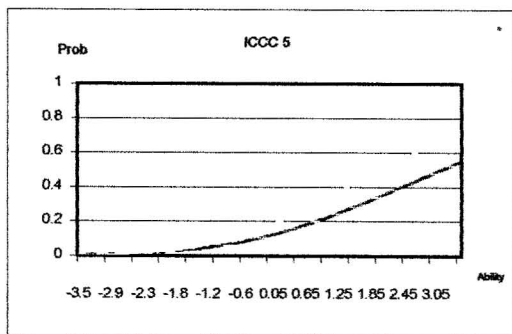
ข้อ 3



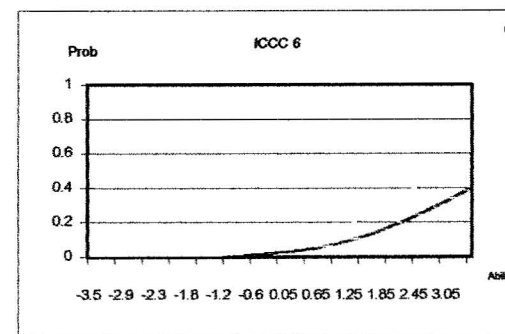
ข้อ 4



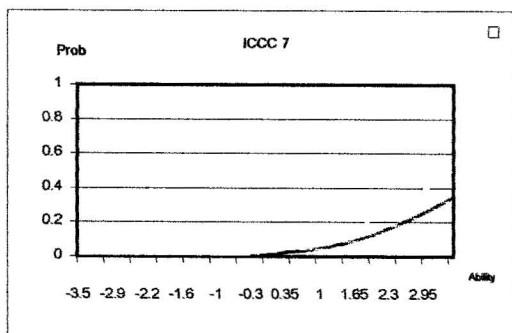
ข้อ 5



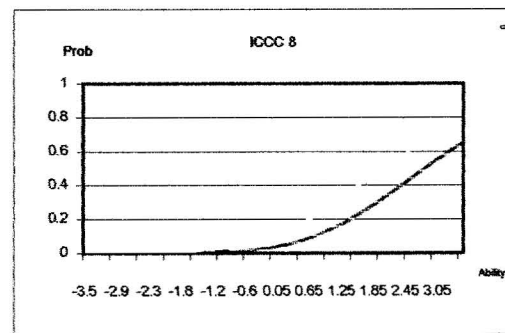
ข้อ 6



ข้อ 7



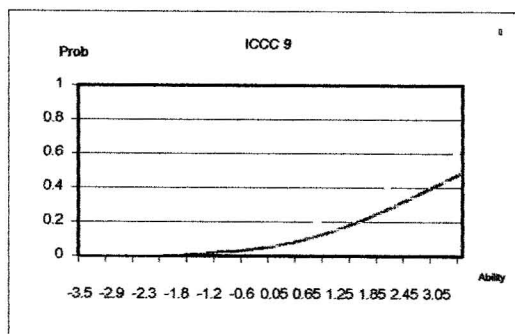
ข้อ 8



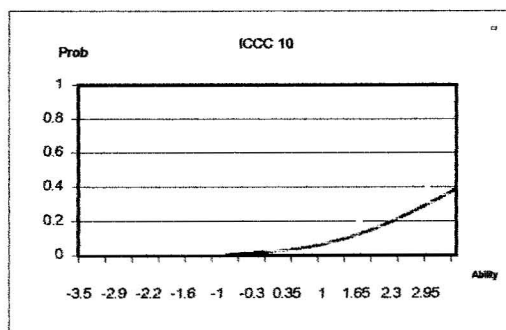
ชาย

หญิง

ข้อ 9

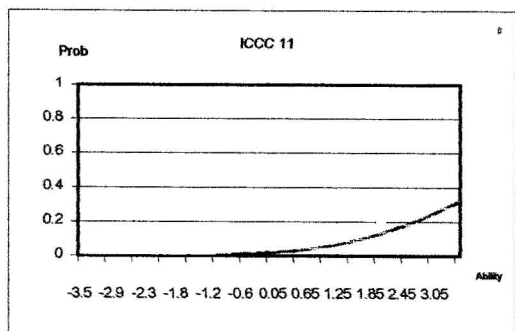


ข้อ 10

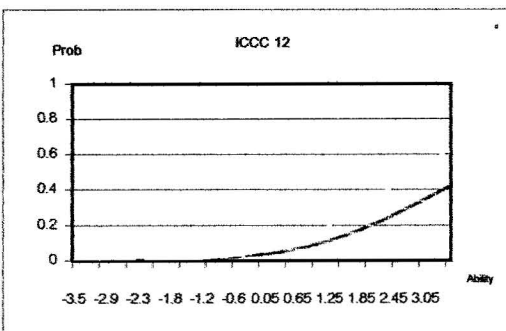


187

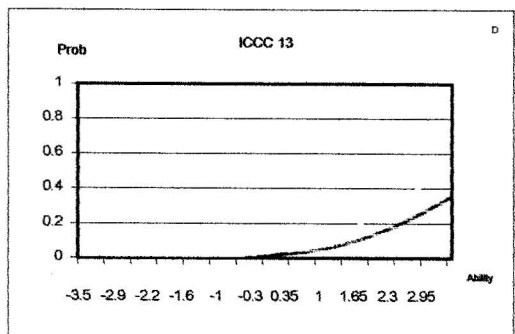
ข้อ 11



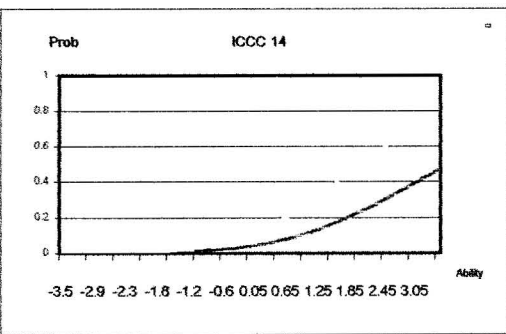
ข้อ 12



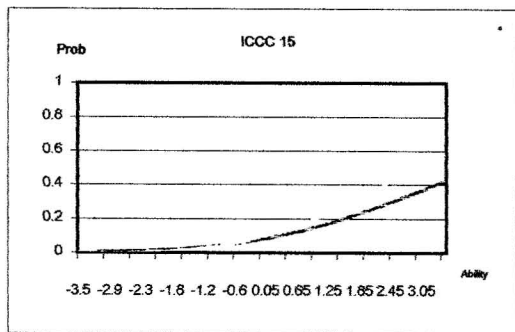
ข้อ 13



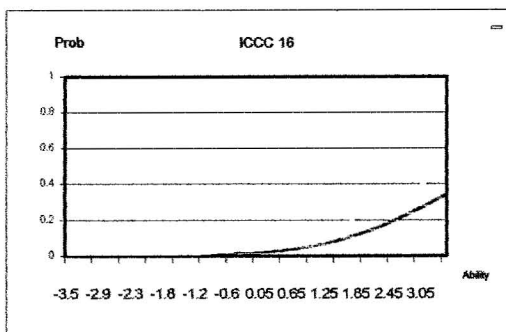
ข้อ 14



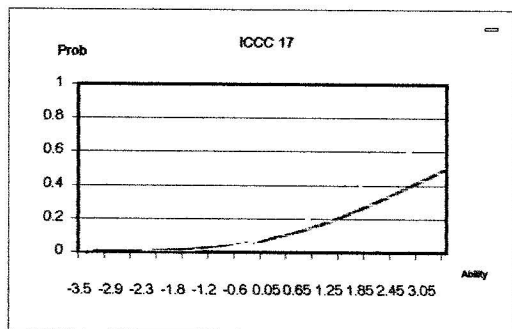
ข้อ 15



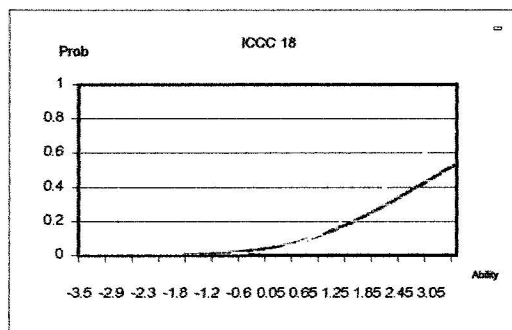
ข้อ 16



ข้อ 17

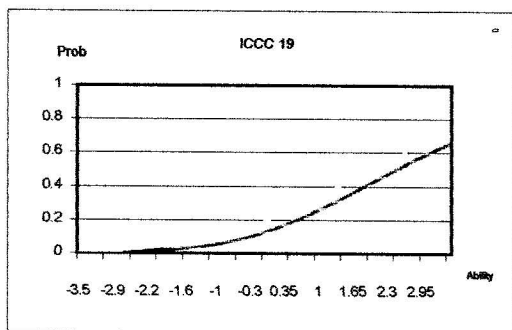


ข้อ 18

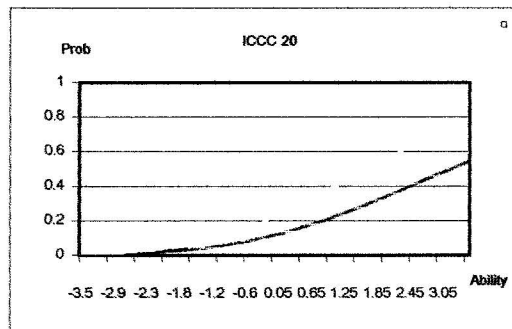


188

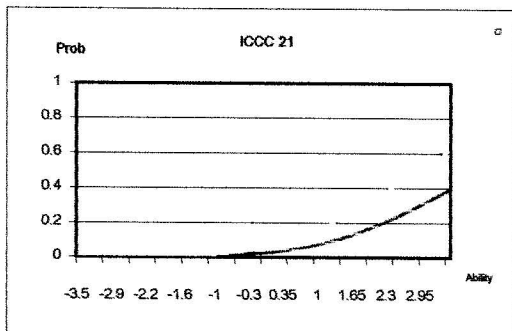
ข้อ 19



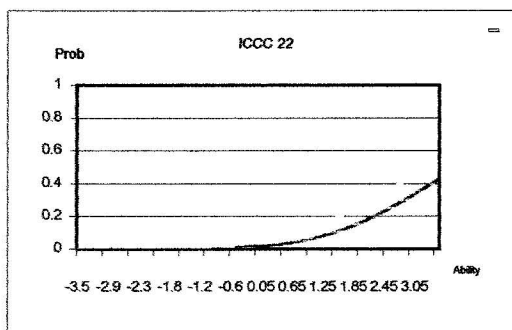
ข้อ 20



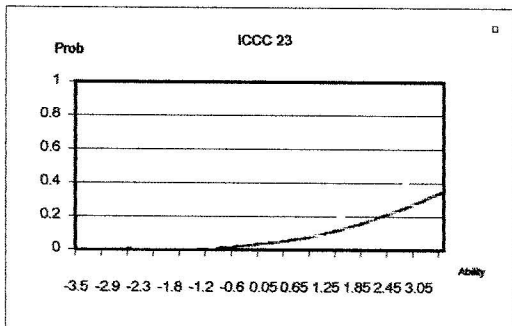
ข้อ 21



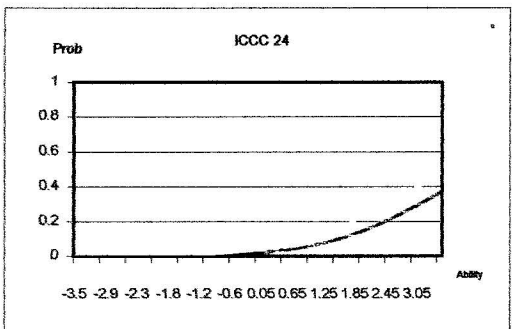
ข้อ 22



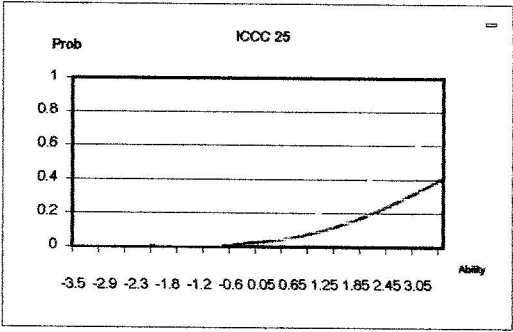
ข้อ 23



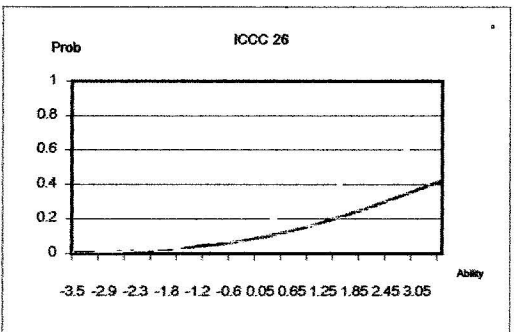
ข้อ 24



ข้อ 25

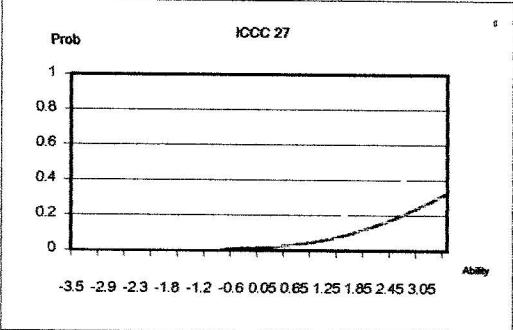


ข้อ 26

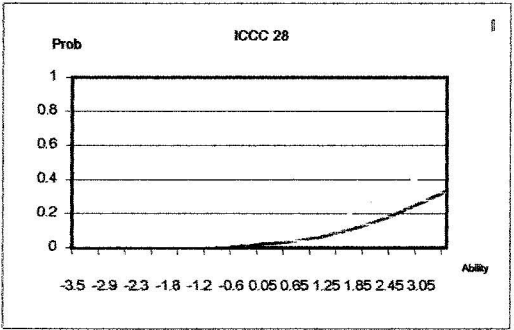


189

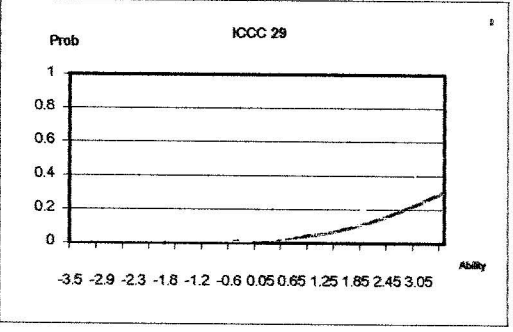
ข้อ 27



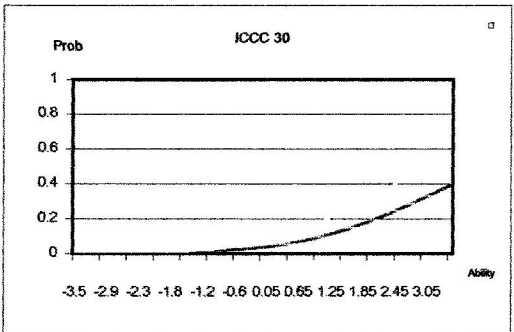
ข้อ 28



ข้อ 29



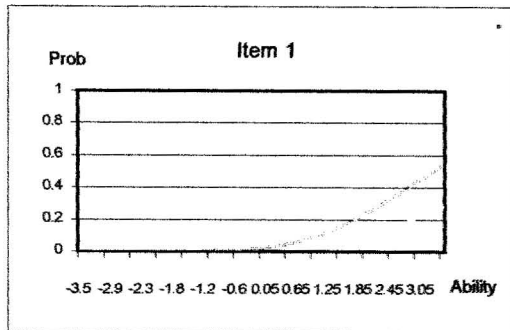
ข้อ 30



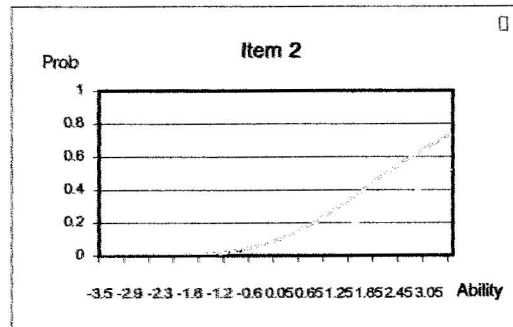
ได้คุณลักษณะของข้อสอบ (ICC) ที่วิเคราะห์ด้วยวิธี GPCM ในแบบทดสอบขนาด 20 ข้อ
เมื่อกลุ่มตัวอย่างเพศชาย ต่อ เพศหญิง เท่ากับ 1,000 คน ต่อ 1,000 คน

190

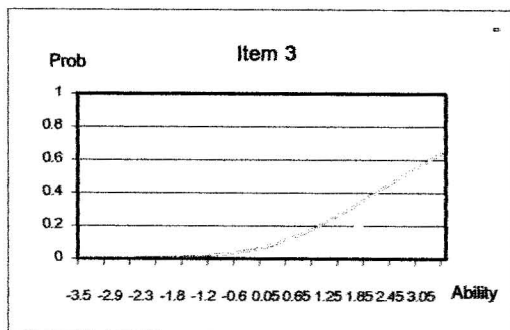
ข้อ 1



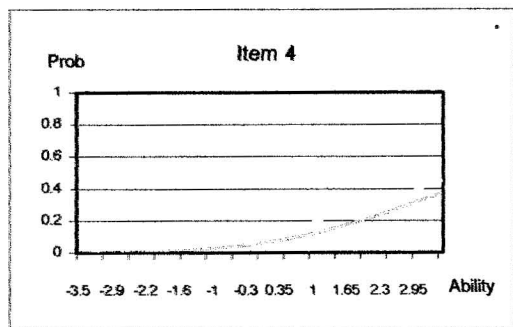
ข้อ 2



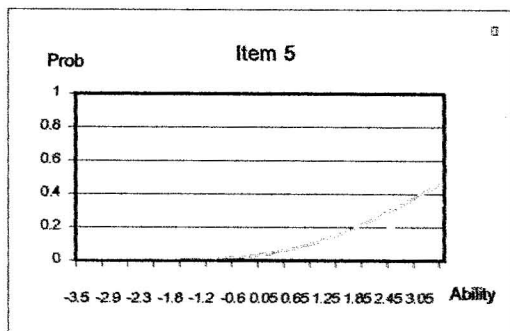
ข้อ 3



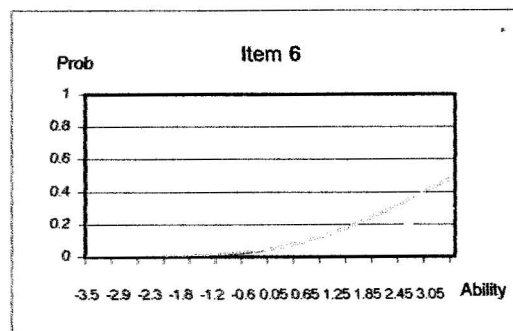
ข้อ 4



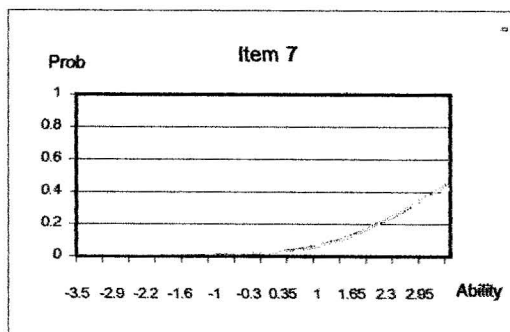
ข้อ 5



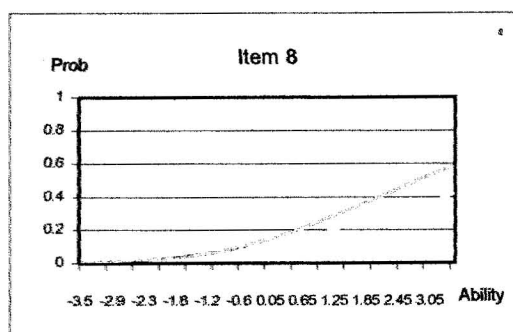
ข้อ 6



ข้อ 7

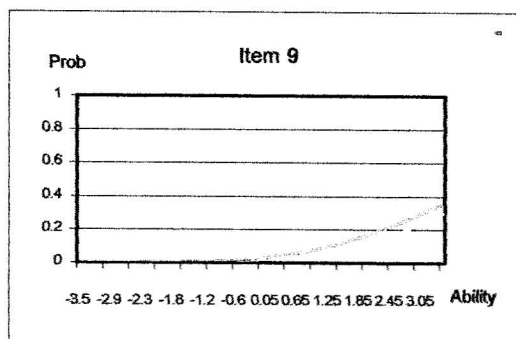


ข้อ 8

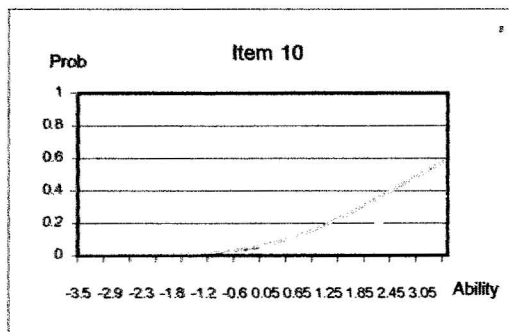


ชาย หญิง

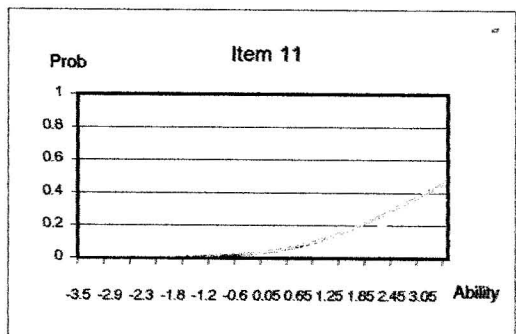
ข้อ 9



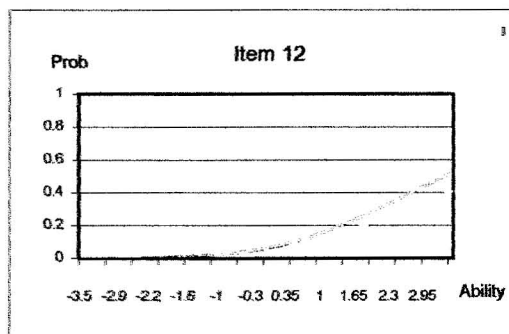
ข้อ 10



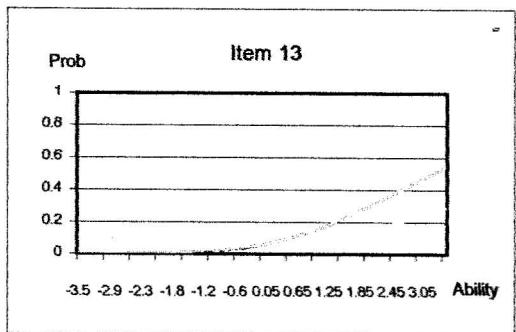
ข้อ 11



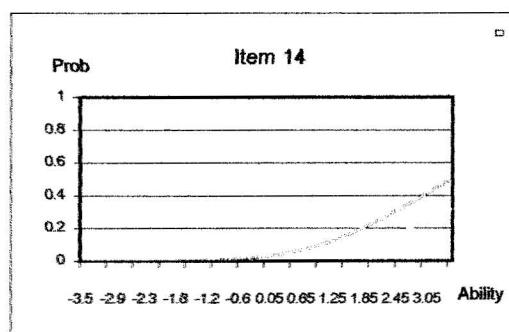
ข้อ 12



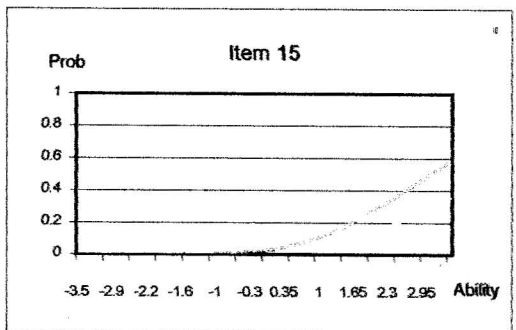
ข้อ 13



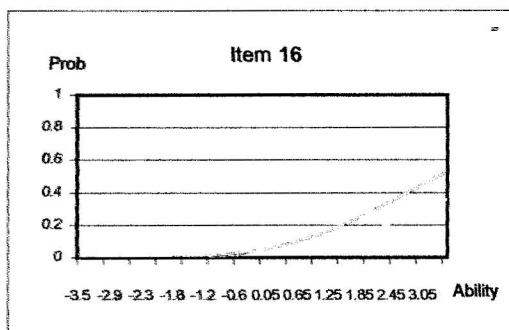
ข้อ 14



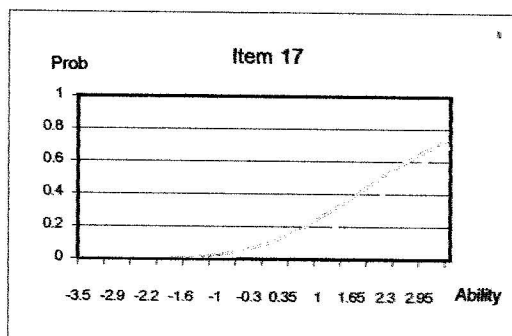
ข้อ 15



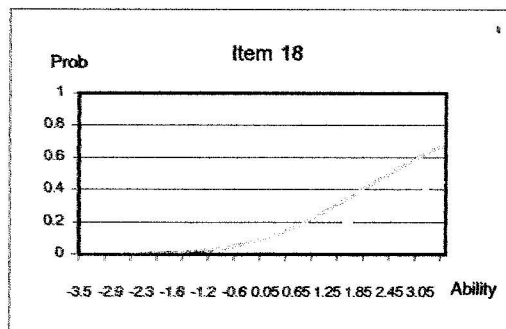
ข้อ 16



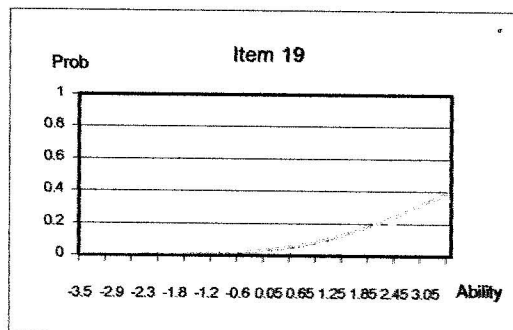
ข้อ 17



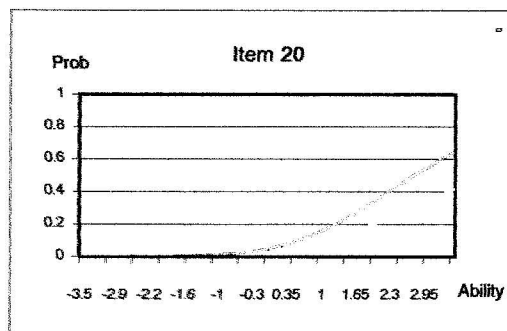
ข้อ 18



ข้อ 19



ข้อ 20



ภาคผนวก ก

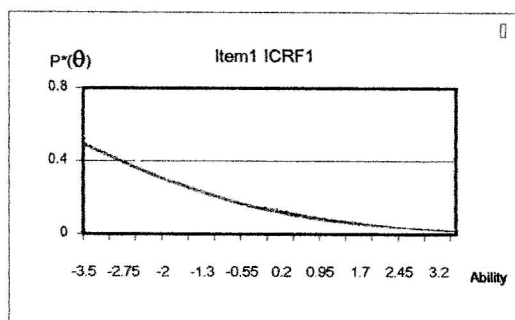
ฟังก์ชันผลการตอบลำดับชั้นข้อสอบ

[Item – Category Response Function : ICRF]

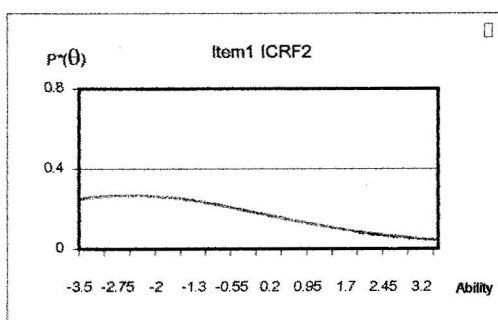
- แบบทดสอบขนาด 40 ข้อ
- แบบทดสอบขนาด 30 ข้อ
- แบบทดสอบขนาด 20 ข้อ

ฟังก์ชันผลการตอบลำดับชั้นข้อสอบ (ICRF) ที่วิเคราะห์ด้วยวิธี GPCM ในแบบทดสอบขนาด 40 ข้อ
เมื่อกลุ่มตัวอย่างเพศชายต่อเพศหญิง เท่ากับ 1,000 คน ต่อ 1,000 คน

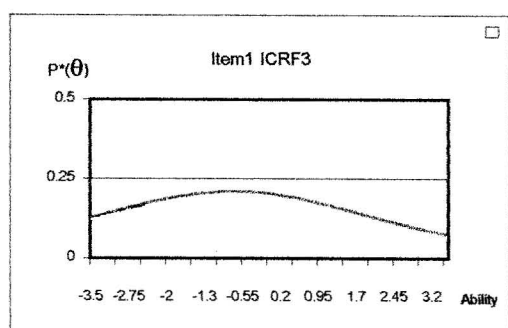
ข้อ 1 ลำดับชั้นที่ 1



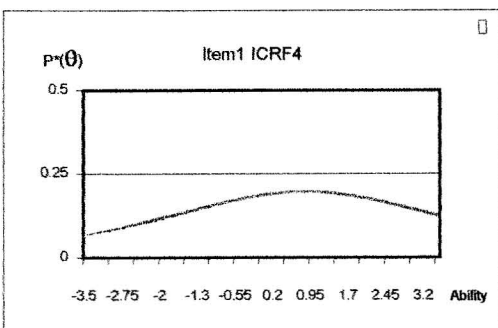
ลำดับชั้นที่ 2



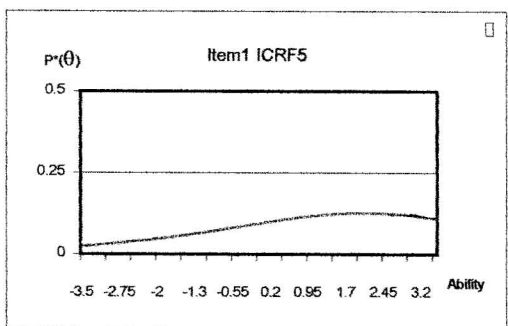
ลำดับชั้นที่ 3



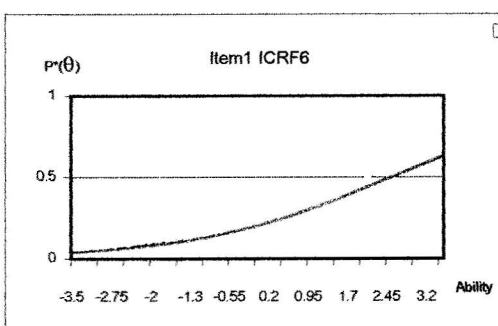
ลำดับชั้นที่ 4



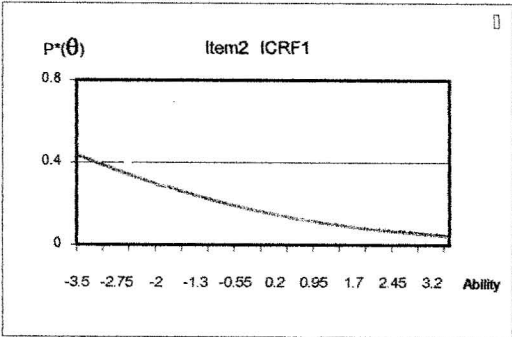
ลำดับชั้นที่ 5



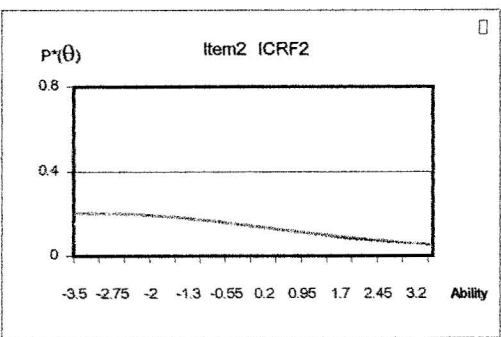
ลำดับชั้นที่ 6



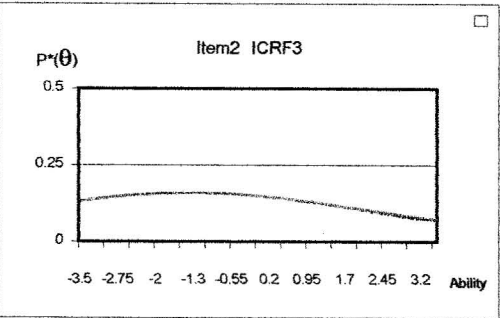
ข้อ 2 ลำดับขั้นที่ 1



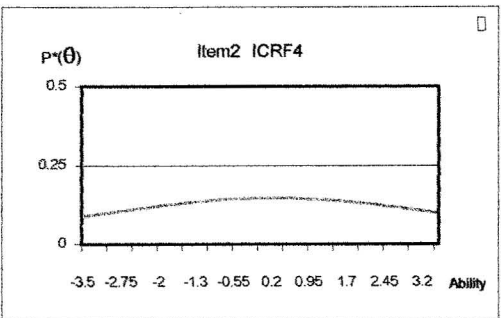
ลำดับขั้นที่ 2



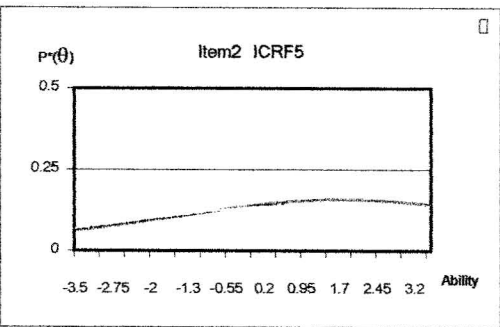
ลำดับขั้นที่ 3



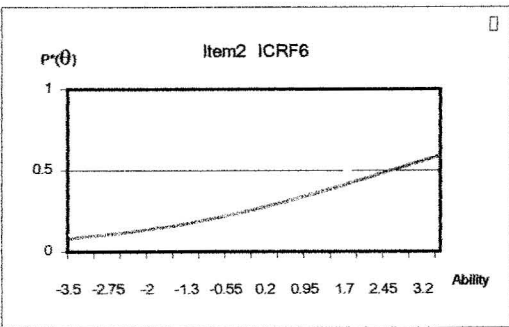
ลำดับขั้นที่ 4



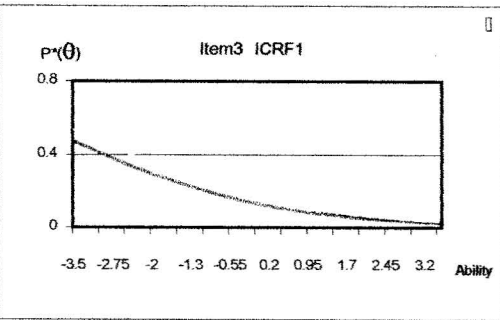
ลำดับขั้นที่ 5



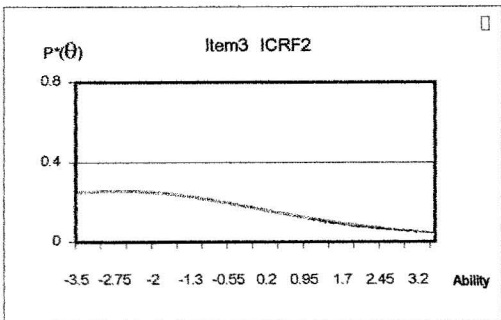
ลำดับขั้นที่ 6



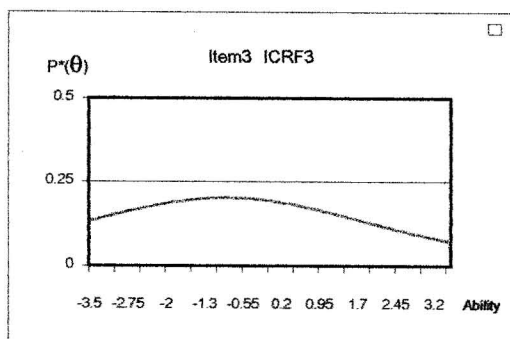
ข้อ 3 ลำดับขั้นที่ 1



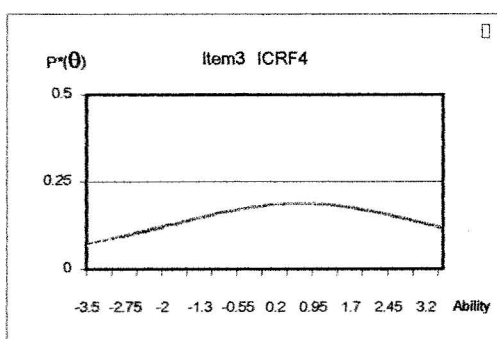
ลำดับขั้นที่ 2



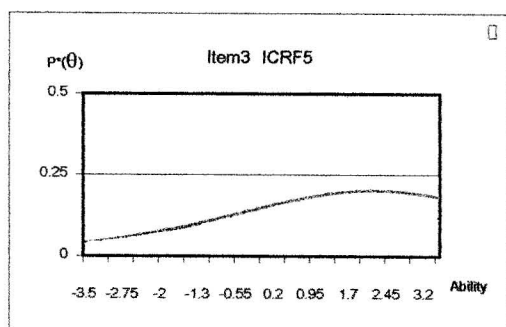
ลำดับขั้นที่ 3



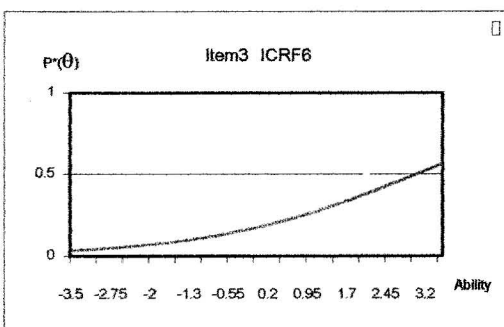
ลำดับขั้นที่ 4



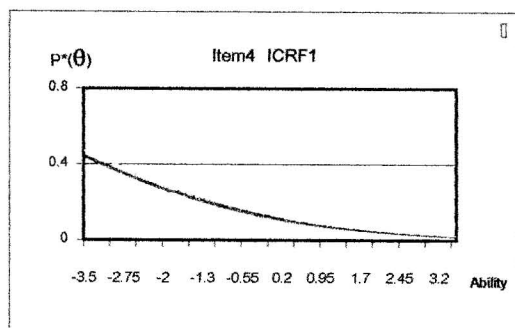
ลำดับขั้นที่ 5



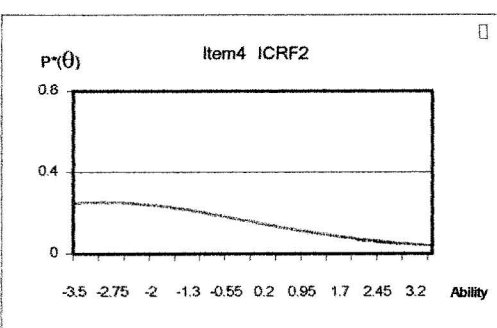
ลำดับขั้นที่ 6



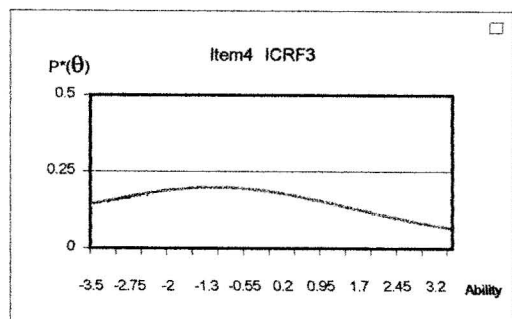
ข้อ 4 ลำดับขั้นที่ 1



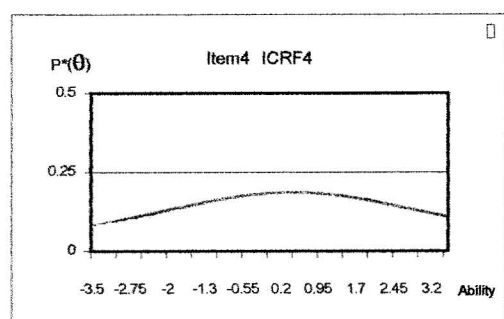
ลำดับขั้นที่ 2



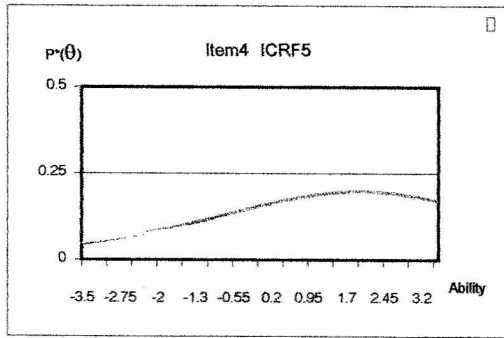
ลำดับขั้นที่ 3



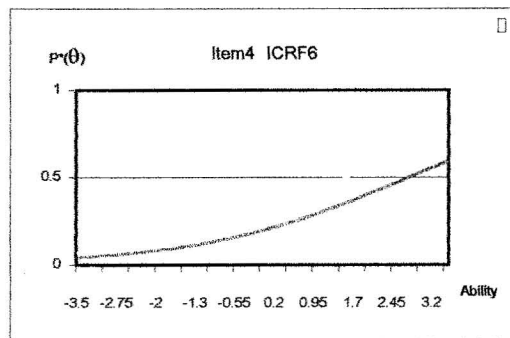
ลำดับขั้นที่ 4



ลำดับชั้นที่ 5

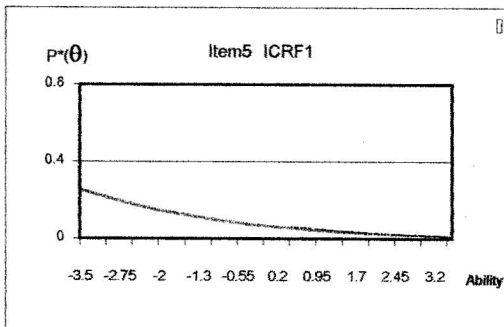


ลำดับชั้นที่ 6

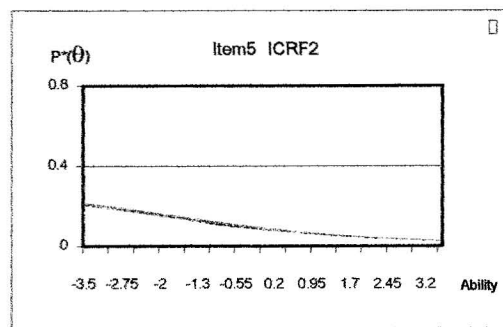


197

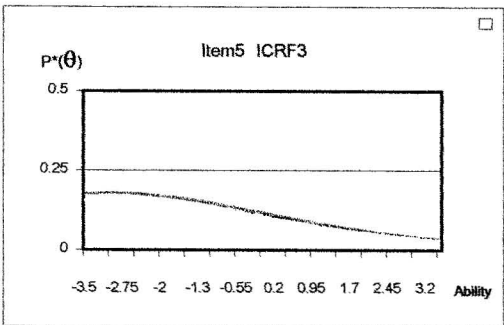
ข้อ 5 ลำดับชั้นที่ 1



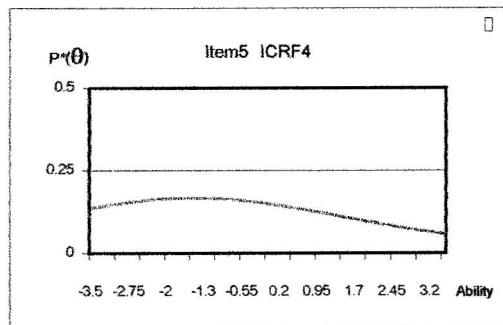
ลำดับชั้นที่ 2



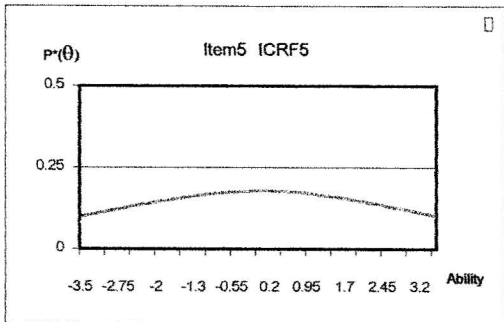
ลำดับชั้นที่ 3



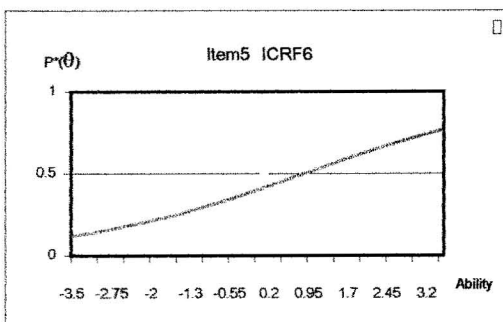
ลำดับชั้นที่ 4



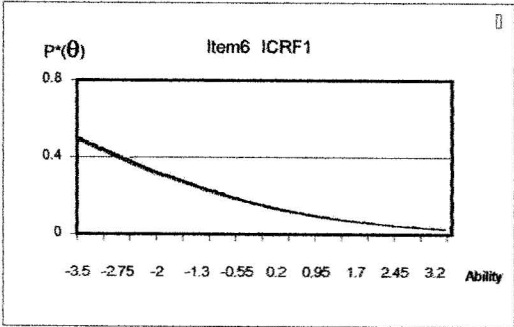
ลำดับชั้นที่ 5



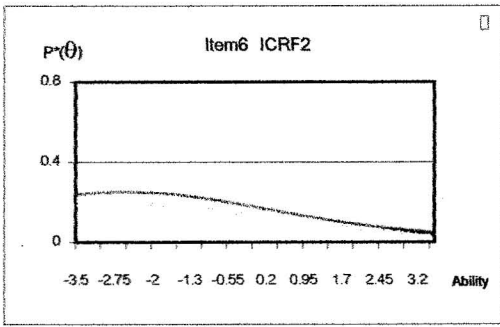
ลำดับชั้นที่ 6



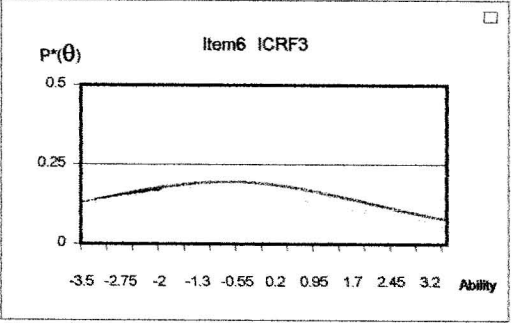
ข้อ 6 ลำดับขั้นที่ 1



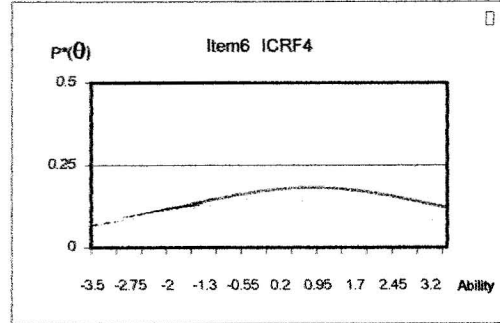
ลำดับขั้นที่ 2



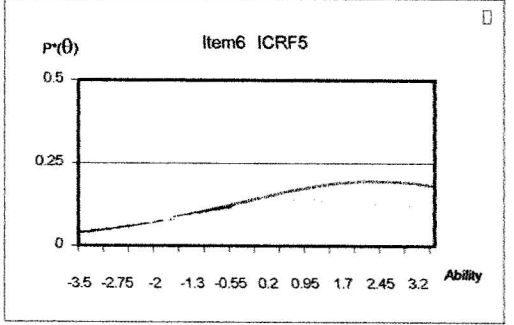
ลำดับขั้นที่ 3



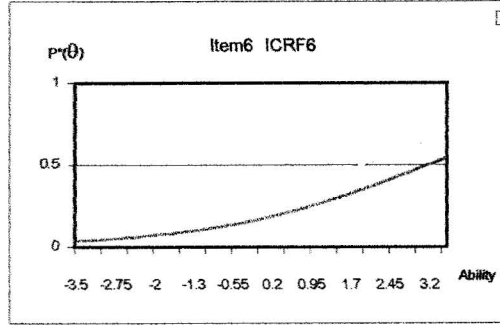
ลำดับขั้นที่ 4



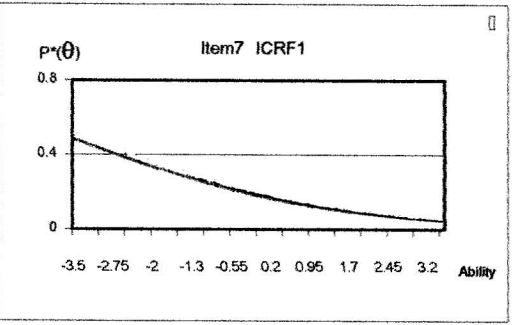
ลำดับขั้นที่ 5



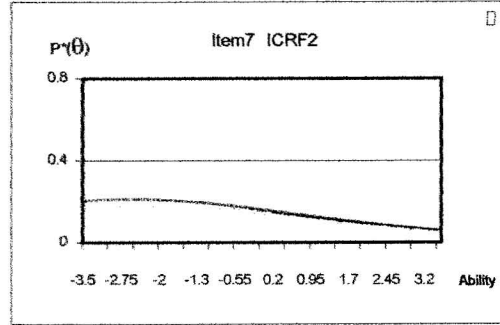
ลำดับขั้นที่ 6



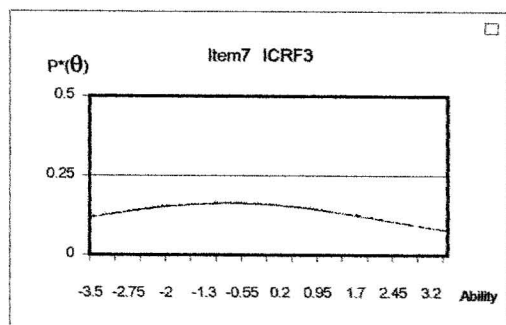
ข้อ 7 ลำดับขั้นที่ 1



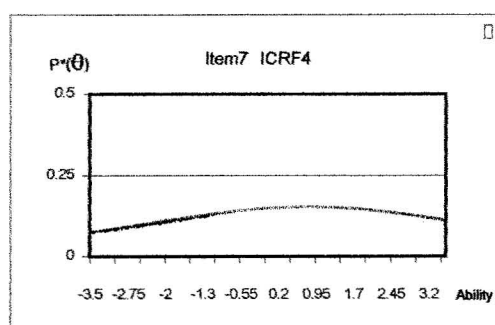
ลำดับขั้นที่ 2



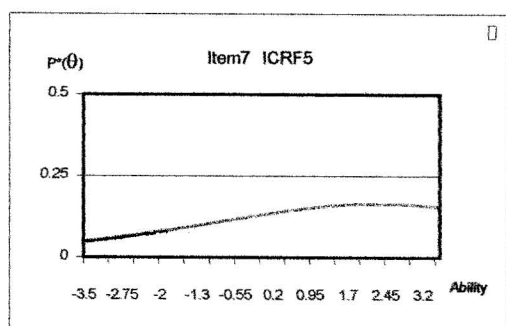
ลำดับชั้นที่ 3



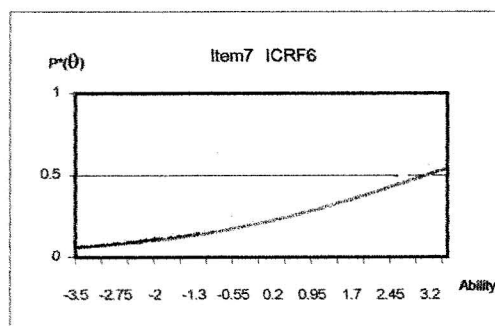
ลำดับชั้นที่ 4



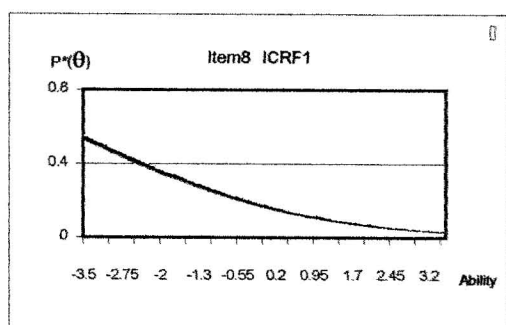
ลำดับชั้นที่ 5



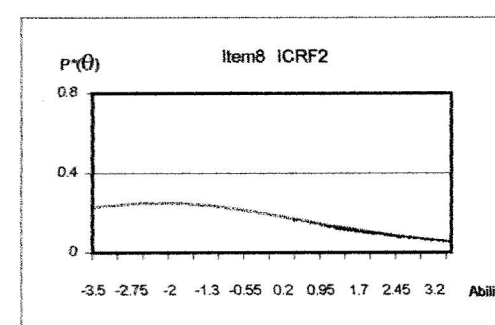
ลำดับชั้นที่ 6



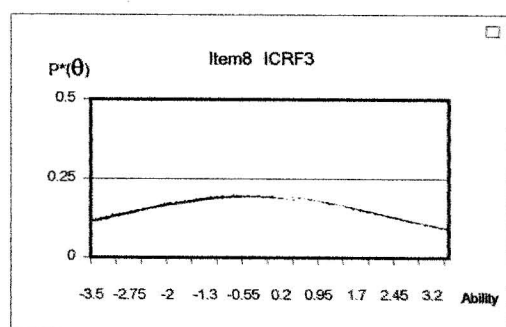
ข้อ 8 ลำดับชั้นที่ 1



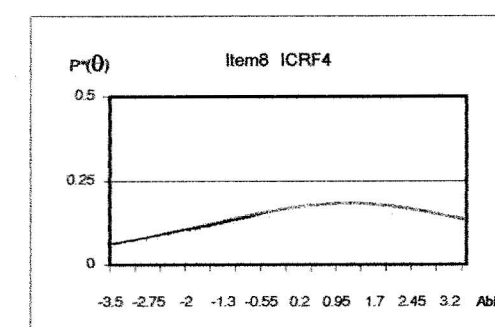
ลำดับชั้นที่ 2



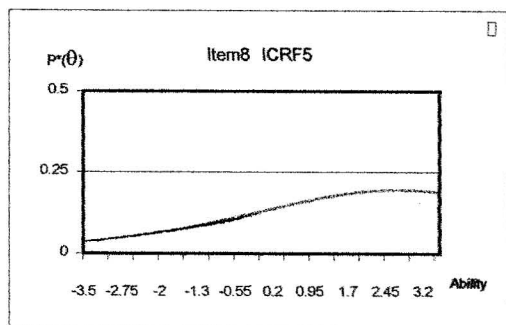
ลำดับชั้นที่ 3



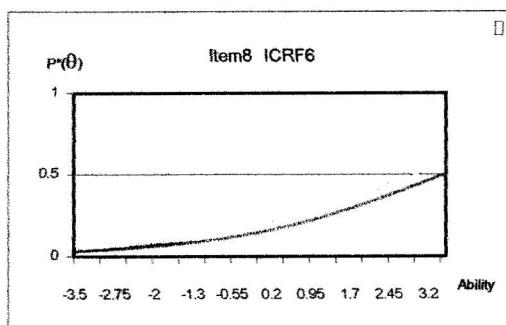
ลำดับชั้นที่ 4



ลำดับขั้นที่ 5

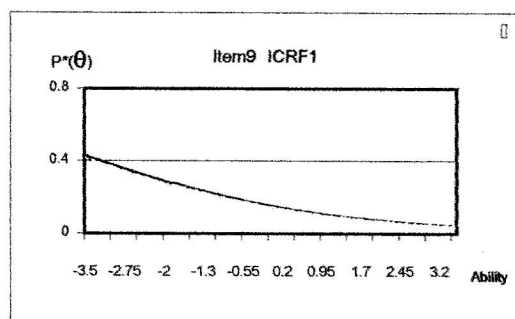


ลำดับขั้นที่ 6

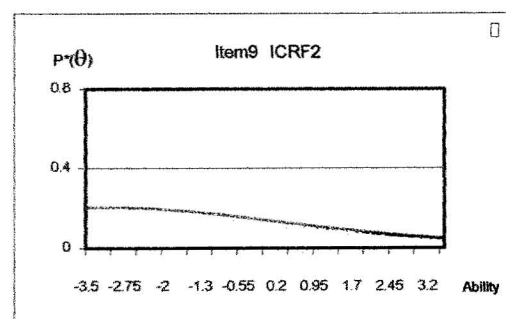


200

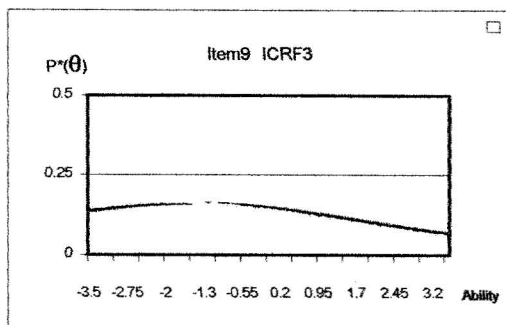
ข้อ 9 ลำดับขั้นที่ 1



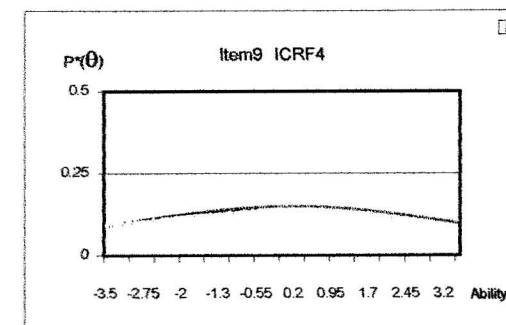
ลำดับขั้นที่ 2



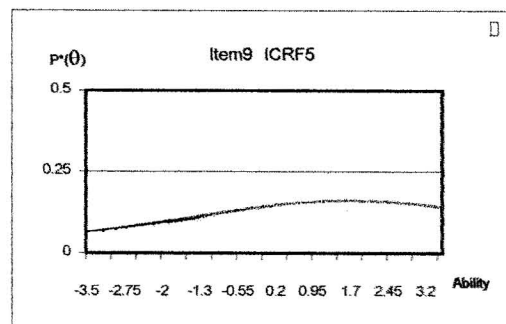
ลำดับขั้นที่ 3



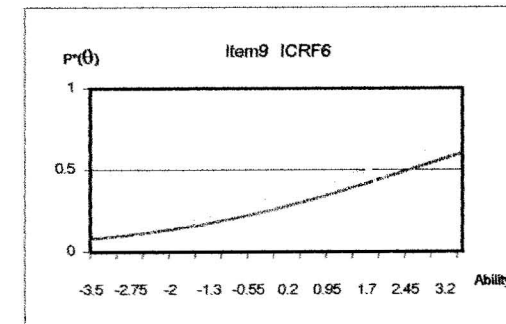
ลำดับขั้นที่ 4



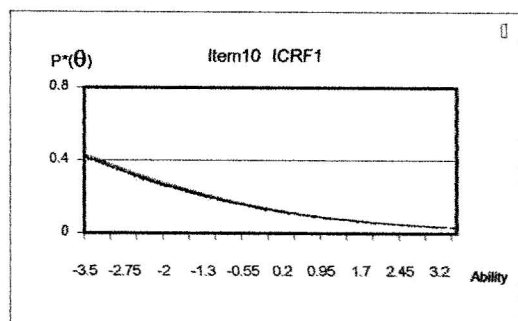
ลำดับขั้นที่ 5



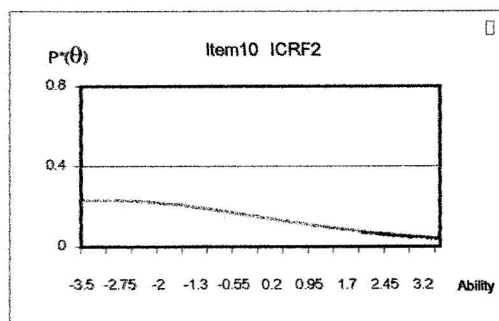
ลำดับขั้นที่ 6



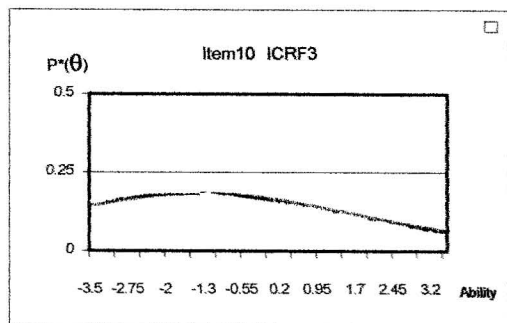
ข้อ 10 ลำดับขั้นที่ 1



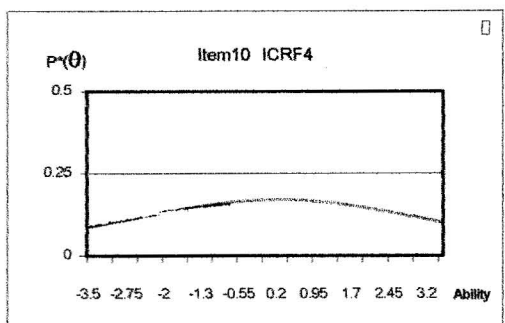
ลำดับขั้นที่ 2



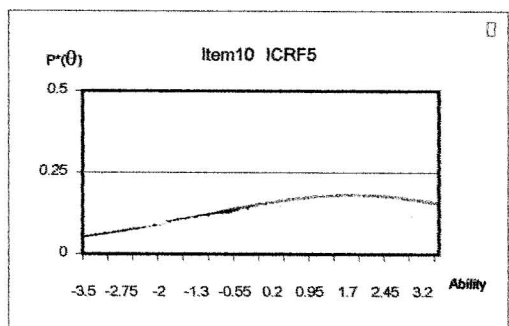
ลำดับขั้นที่ 3



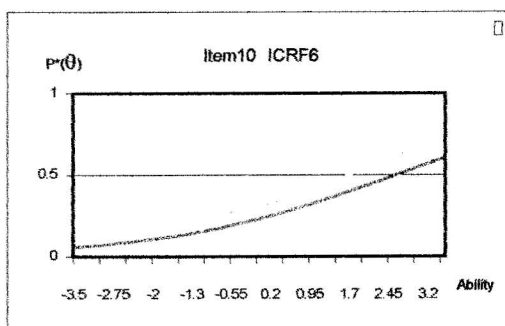
ลำดับขั้นที่ 4



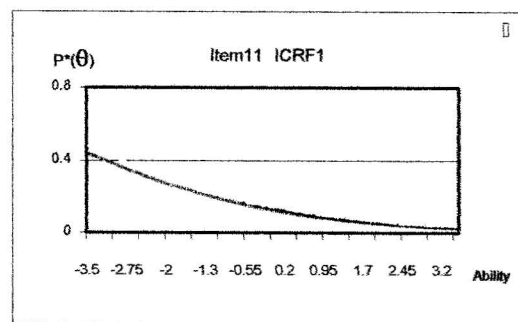
ลำดับขั้นที่ 5



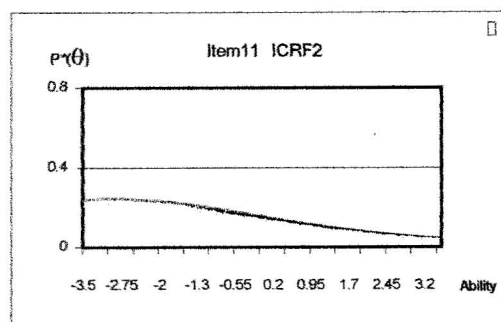
ลำดับขั้นที่ 6



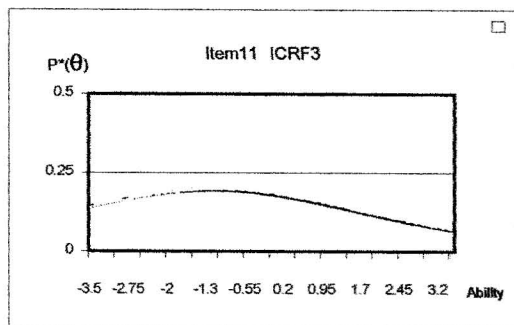
ข้อ 11 ลำดับขั้นที่ 1



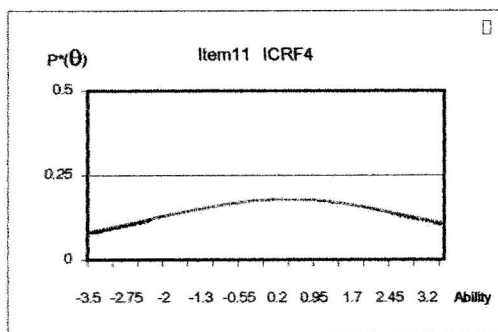
ลำดับขั้นที่ 2



ลำดับขั้นที่ 3

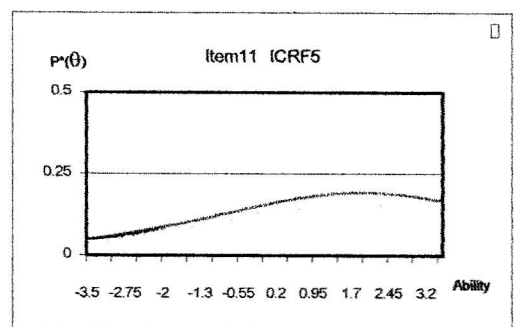


ลำดับขั้นที่ 4

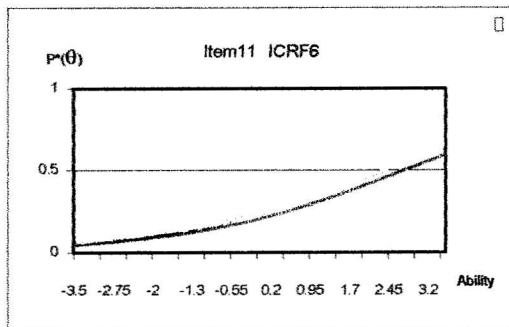


202

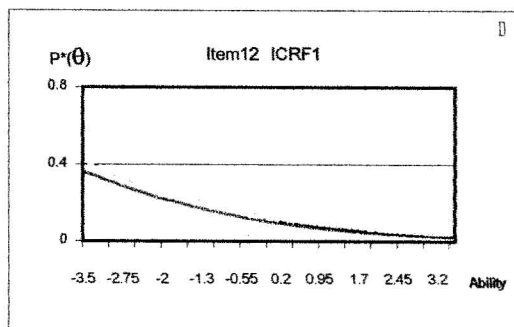
ลำดับขั้นที่ 5



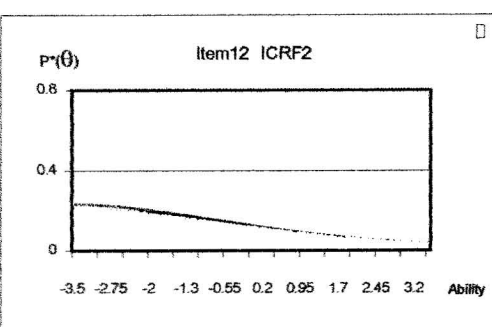
ลำดับขั้นที่ 6



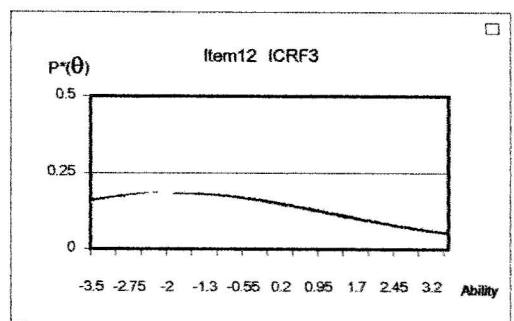
ข้อ 12 ลำดับขั้นที่ 1



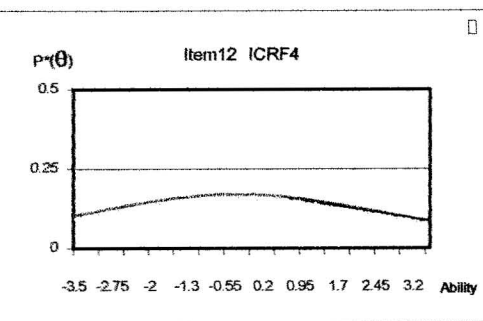
ลำดับขั้นที่ 2

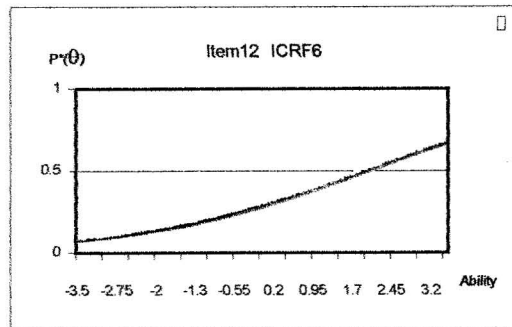
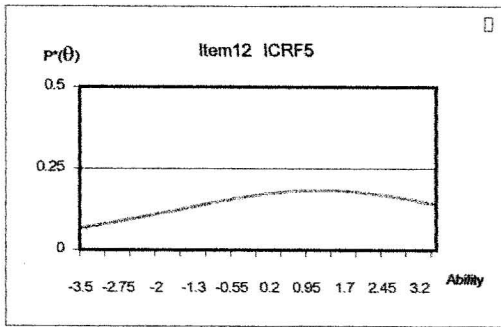


ลำดับขั้นที่ 3



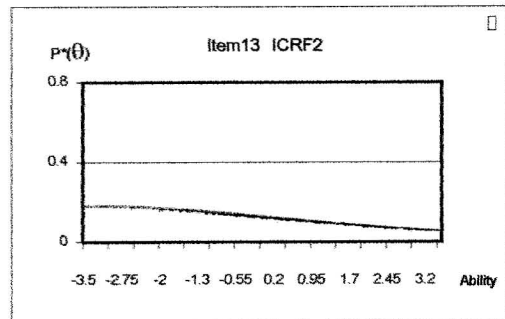
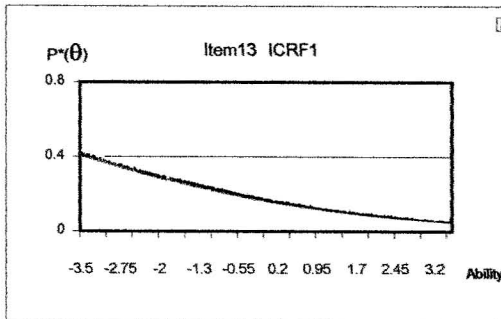
ลำดับขั้นที่ 4





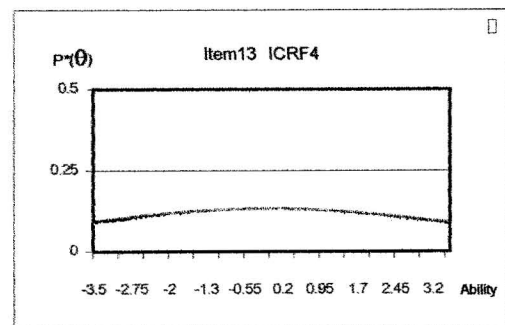
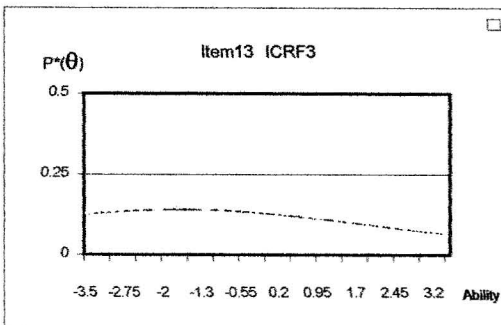
ข้อ 13 ลำดับขั้นที่ 1

ลำดับขั้นที่ 2



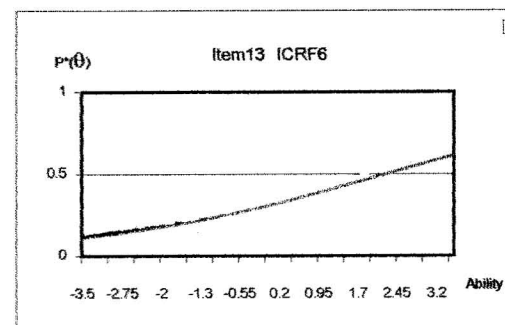
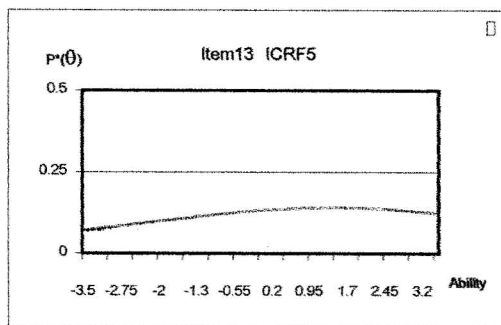
ลำดับขั้นที่ 3

ลำดับขั้นที่ 4

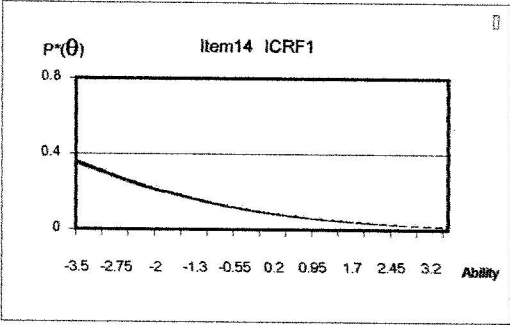


ลำดับขั้นที่ 5

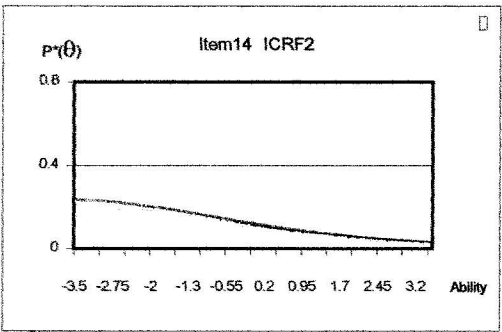
ลำดับขั้นที่ 6



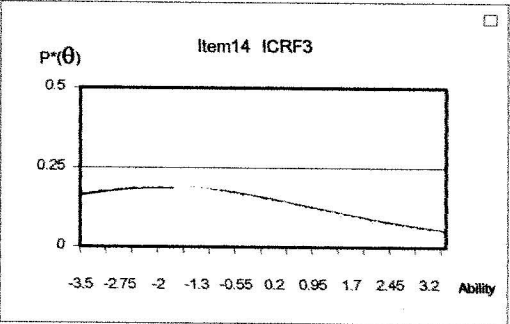
ข้อ 14 ลำดับขั้นที่ 1



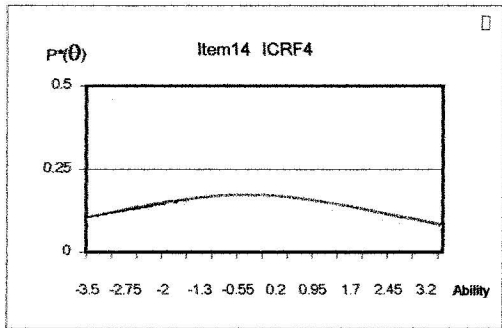
ลำดับขั้นที่ 2



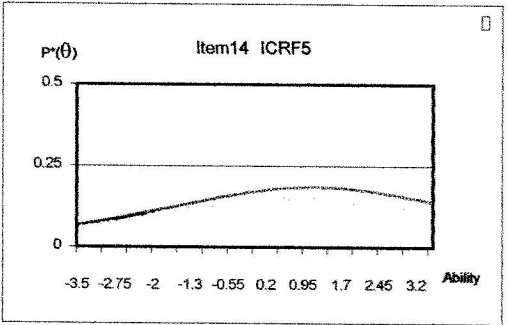
ลำดับขั้นที่ 3



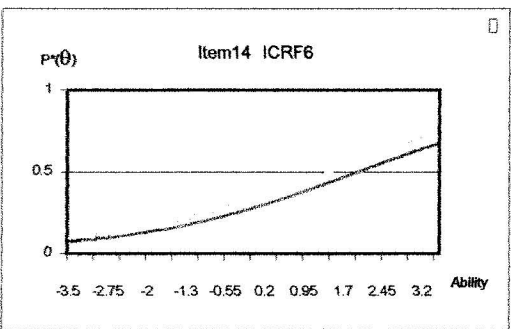
ลำดับขั้นที่ 4



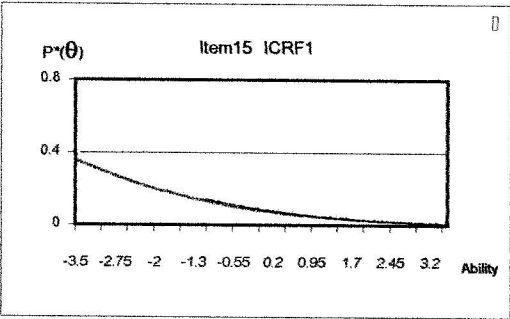
ลำดับขั้นที่ 5



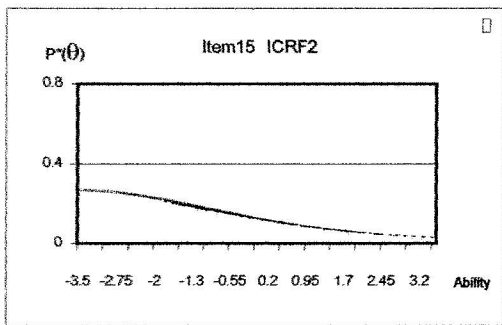
ลำดับขั้นที่ 6



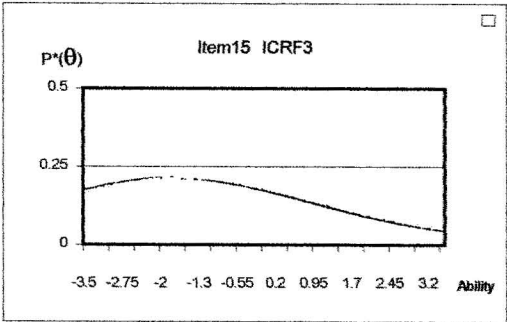
ข้อ 15 ลำดับขั้นที่ 1



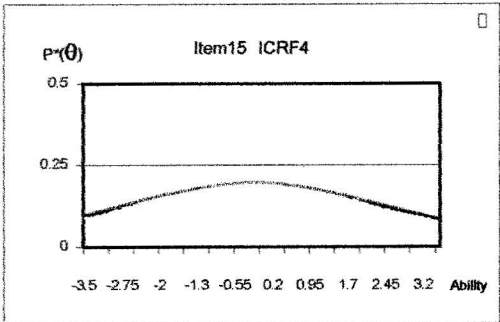
ลำดับขั้นที่ 2



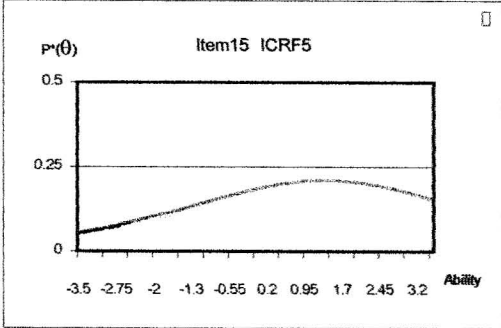
ลำดับชั้นที่ 3



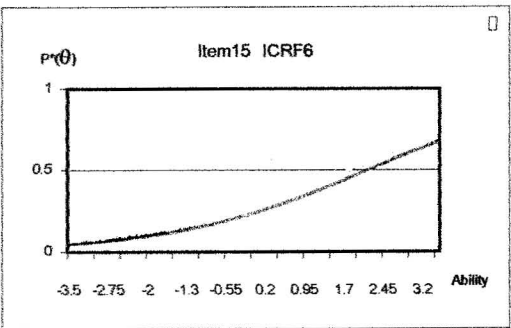
ลำดับชั้นที่ 4



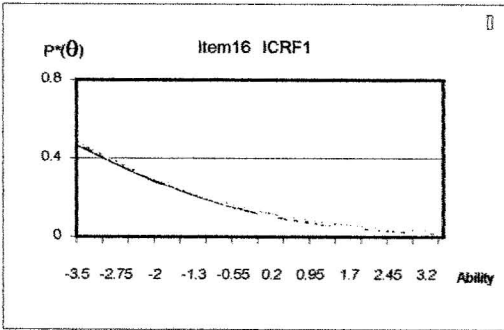
ลำดับชั้นที่ 5



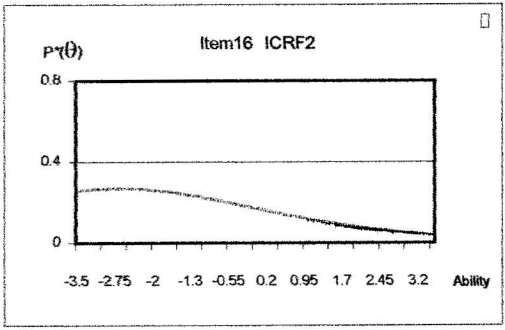
ลำดับชั้นที่ 6



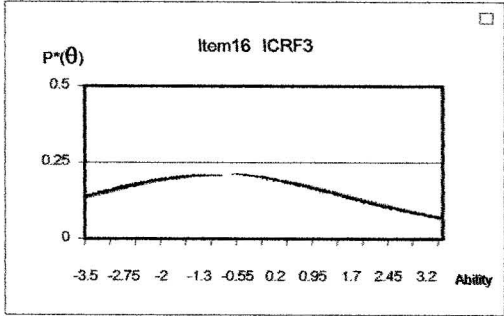
ข้อ 16 ลำดับชั้นที่ 1



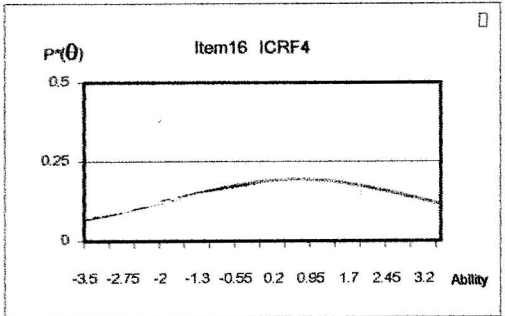
ลำดับชั้นที่ 2



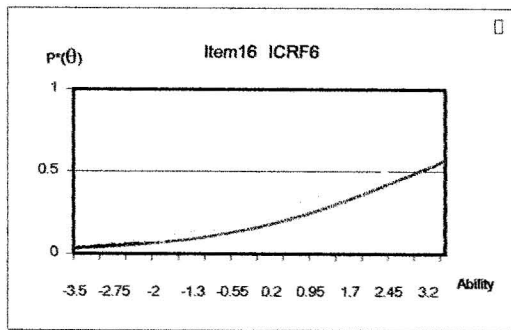
ลำดับชั้นที่ 3



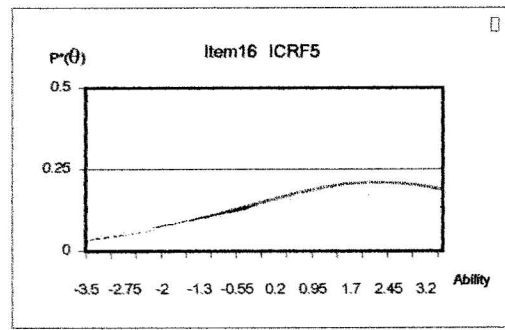
ลำดับชั้นที่ 4



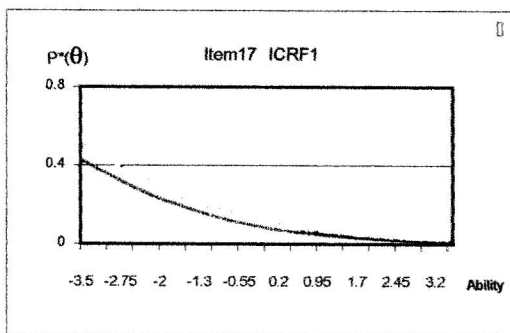
ลำดับขั้นที่ 5



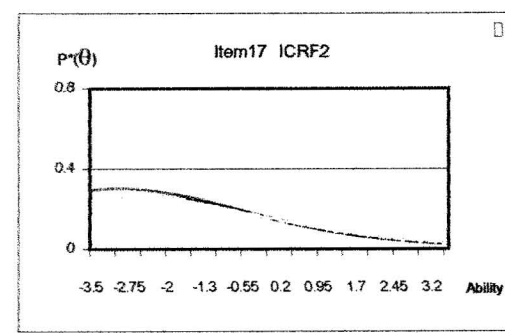
ลำดับขั้นที่ 6



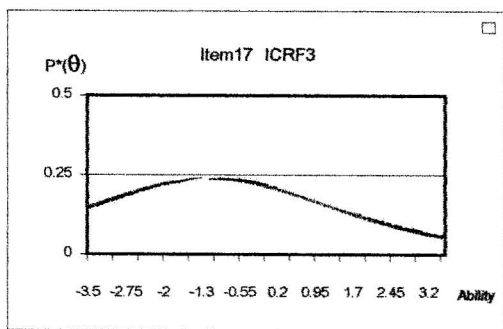
ข้อ 17 ลำดับขั้นที่ 1



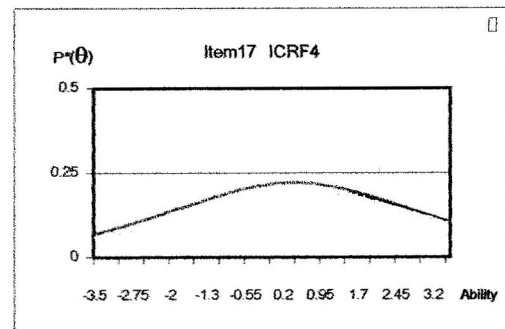
ลำดับขั้นที่ 2



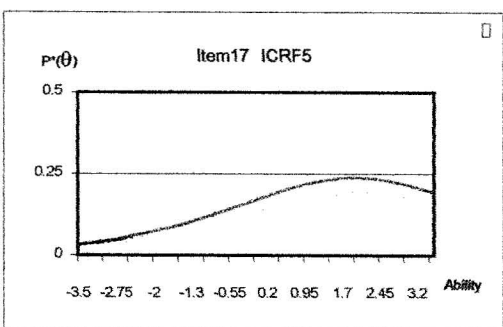
ลำดับขั้นที่ 3



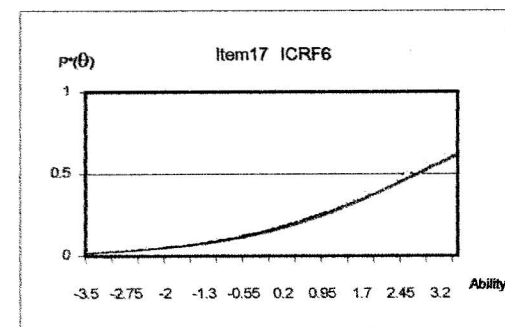
ลำดับขั้นที่ 4



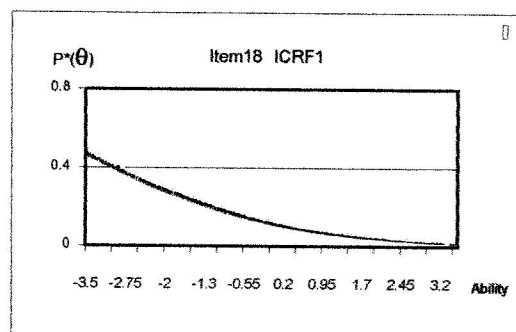
ลำดับขั้นที่ 5



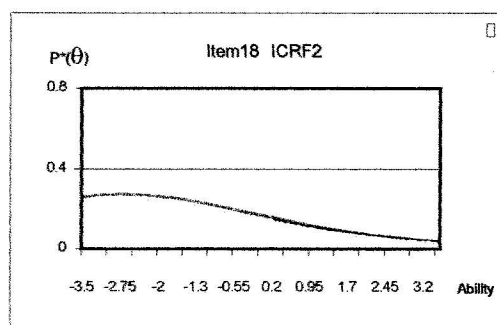
ลำดับขั้นที่ 6



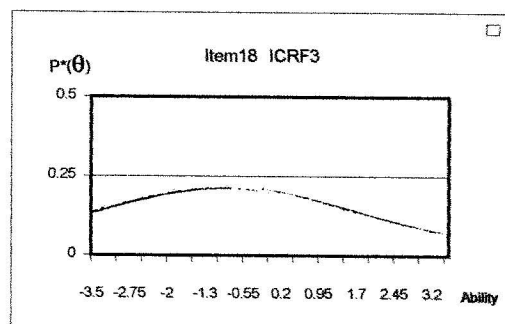
ข้อ 18 ลำดับขั้นที่ 1



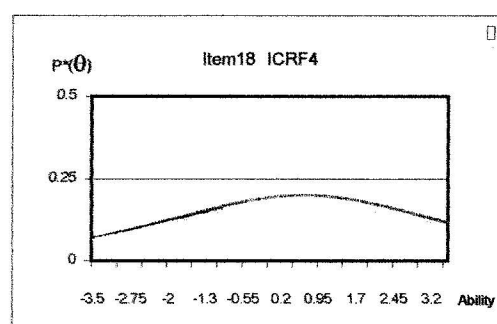
ลำดับขั้นที่ 2



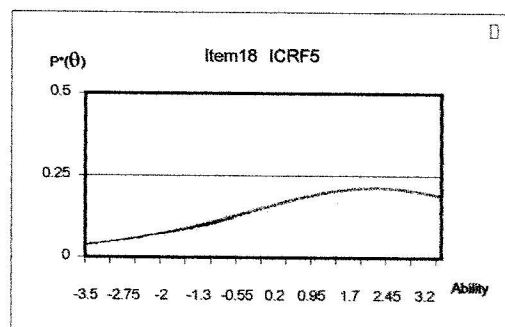
ลำดับขั้นที่ 3



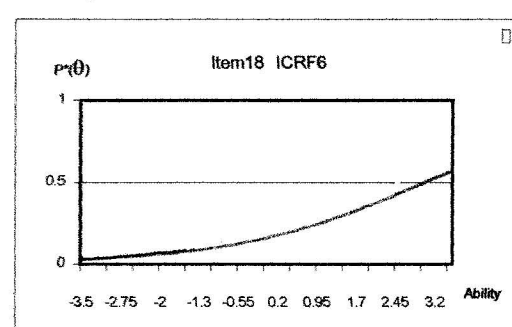
ลำดับขั้นที่ 4



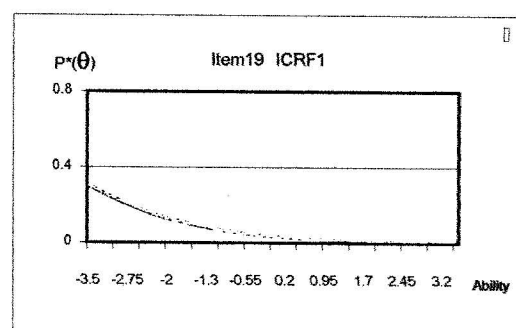
ลำดับขั้นที่ 5



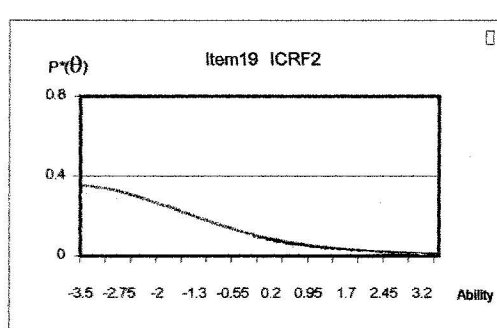
ลำดับขั้นที่ 6



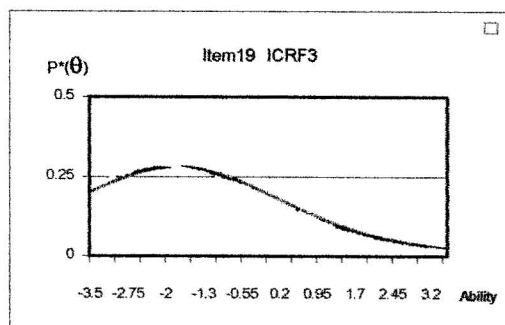
ข้อ 19 ลำดับขั้นที่ 1



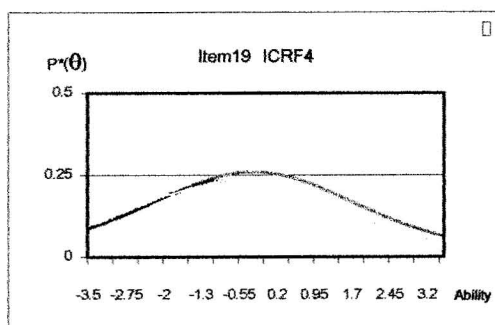
ลำดับขั้นที่ 2



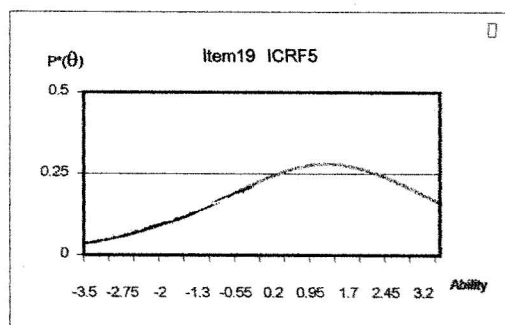
ลำดับขั้นที่ 3



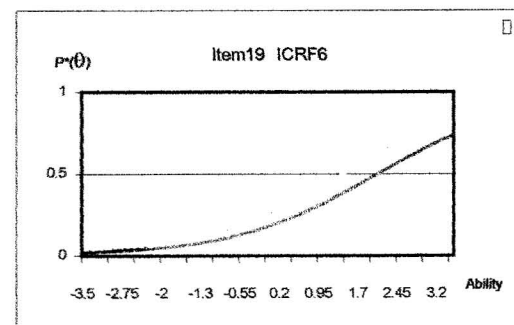
ลำดับขั้นที่ 4



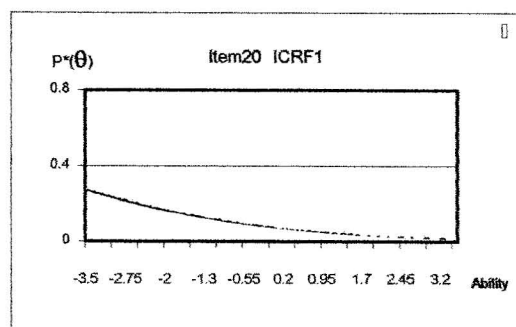
ลำดับขั้นที่ 5



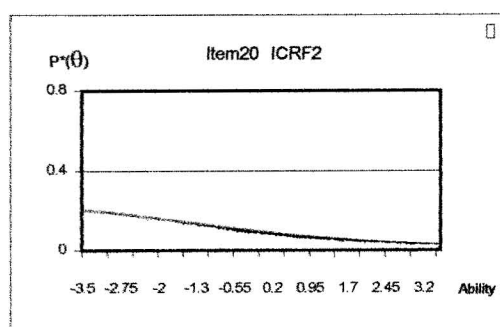
ลำดับขั้นที่ 6



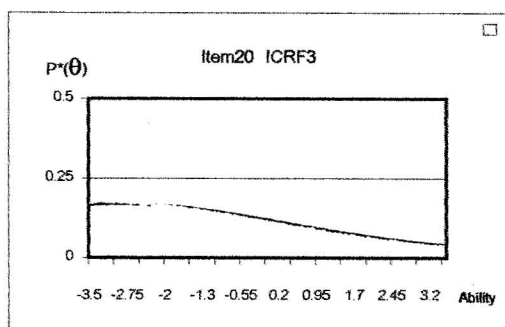
ข้อ 20 ลำดับขั้นที่ 1



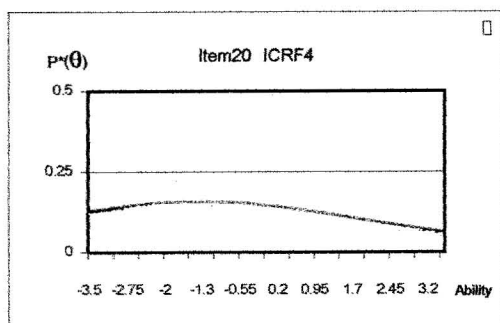
ลำดับขั้นที่ 2



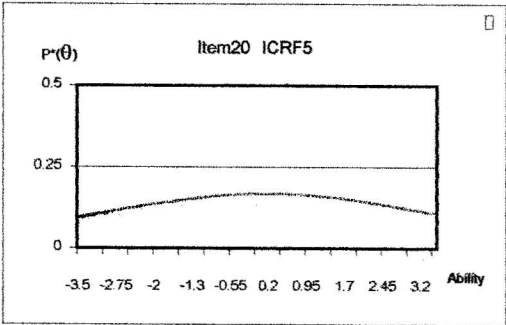
ลำดับขั้นที่ 3



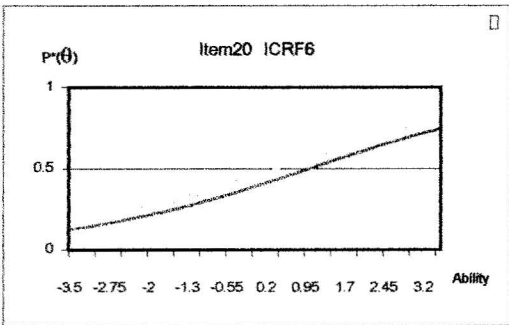
ลำดับขั้นที่ 4



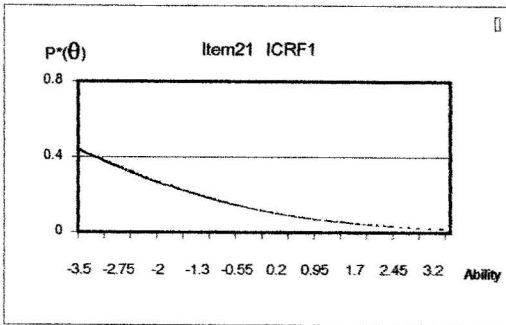
ลำดับชั้นที่ 5



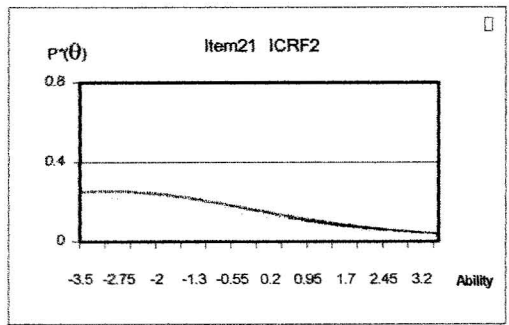
ลำดับชั้นที่ 6



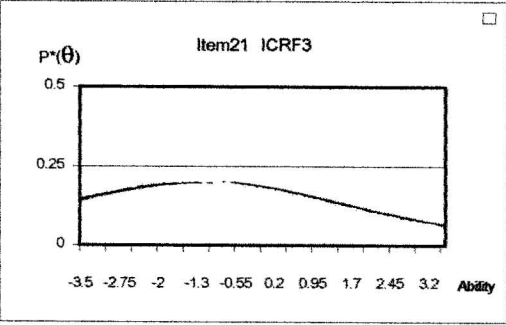
ข้อ 21 ลำดับชั้นที่ 1



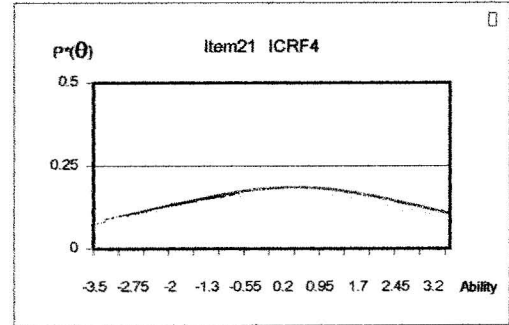
ลำดับชั้นที่ 2



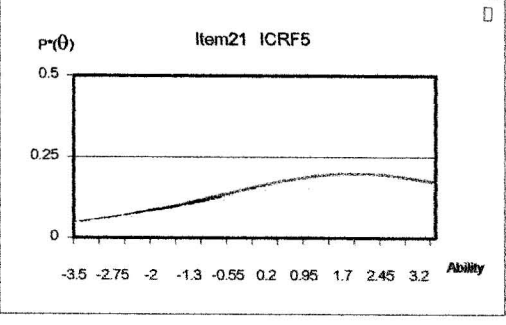
ลำดับชั้นที่ 3



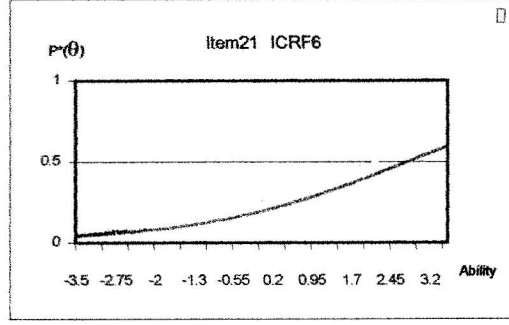
ลำดับชั้นที่ 4



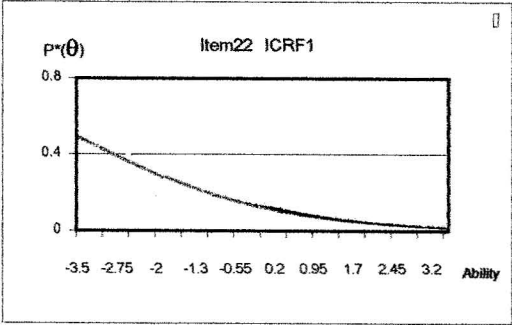
ลำดับชั้นที่ 5



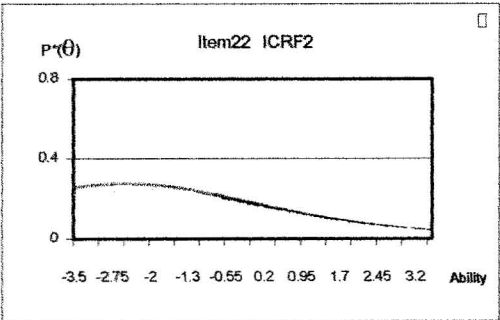
ลำดับชั้นที่ 6



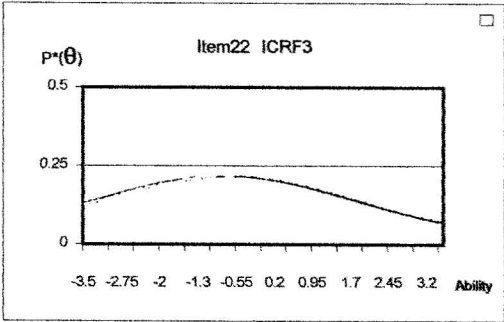
ข้อ 22 ลำดับขั้นที่ 1



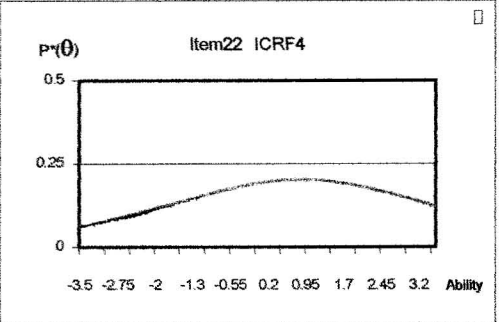
ลำดับขั้นที่ 2



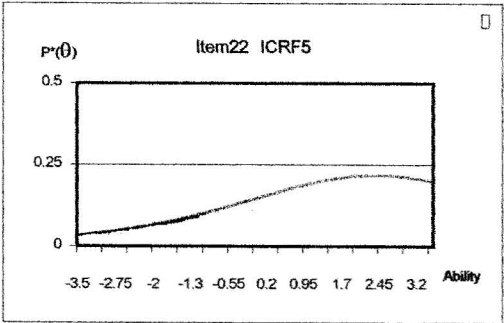
ลำดับขั้นที่ 3



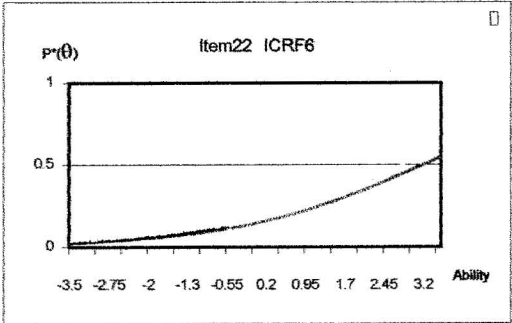
ลำดับขั้นที่ 4



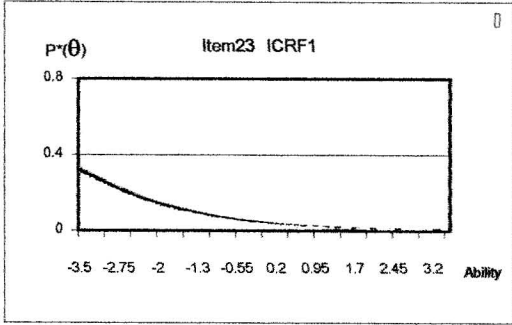
ลำดับขั้นที่ 5



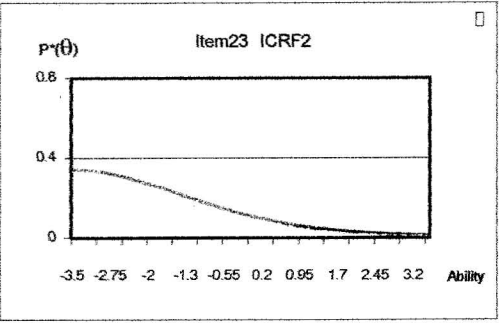
ลำดับขั้นที่ 6



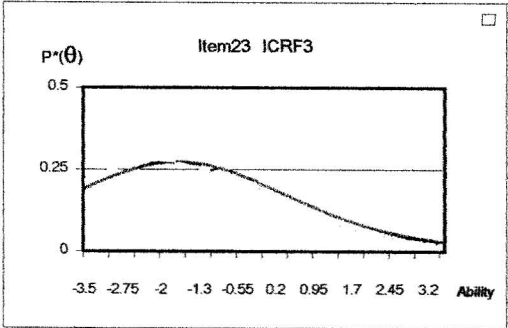
ข้อ 23 ลำดับขั้นที่ 1



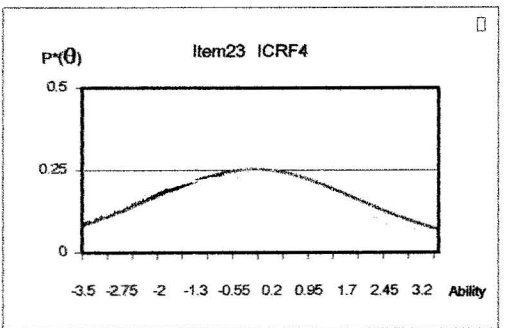
ลำดับขั้นที่ 2



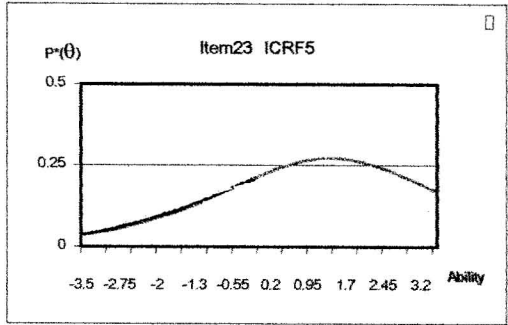
ลำดับชั้นที่ 3



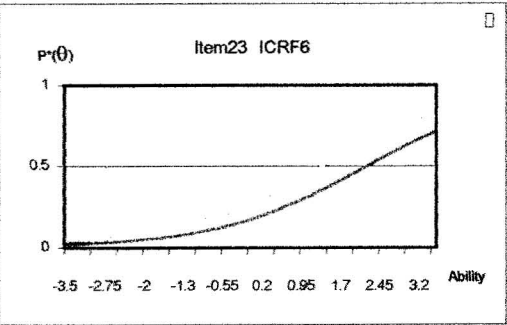
ลำดับชั้นที่ 4



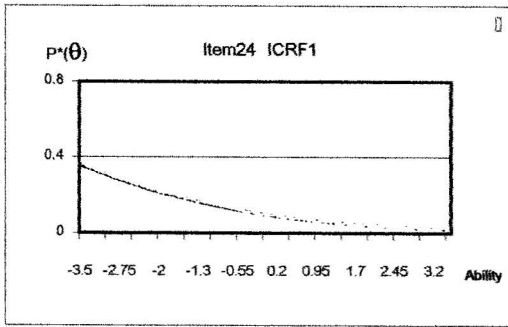
ลำดับชั้นที่ 5



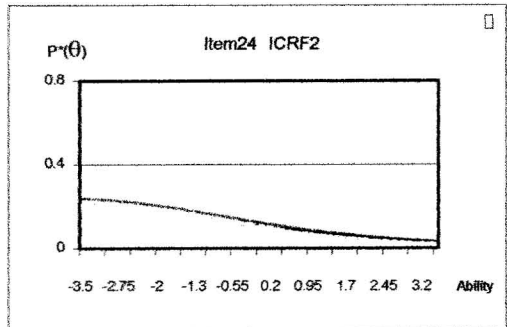
ลำดับชั้นที่ 6



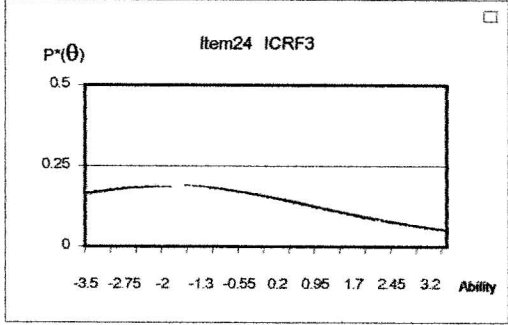
ข้อ 24 ลำดับชั้นที่ 1



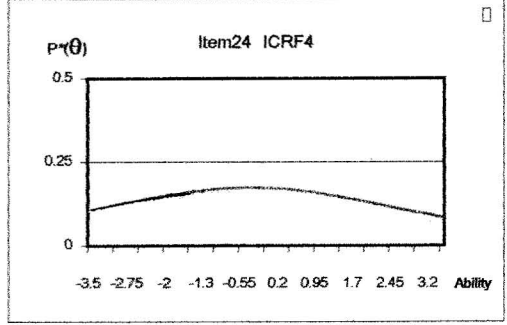
ลำดับชั้นที่ 2



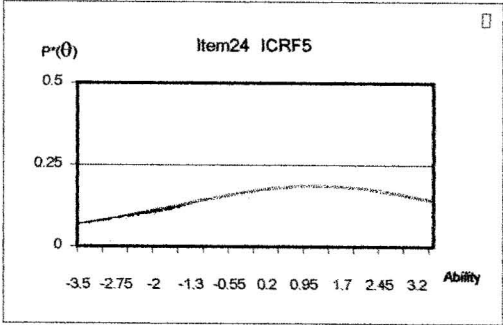
ลำดับชั้นที่ 3



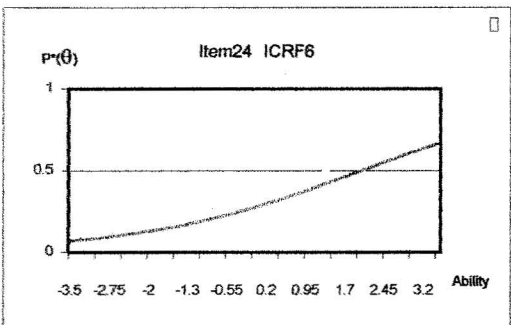
ลำดับชั้นที่ 4



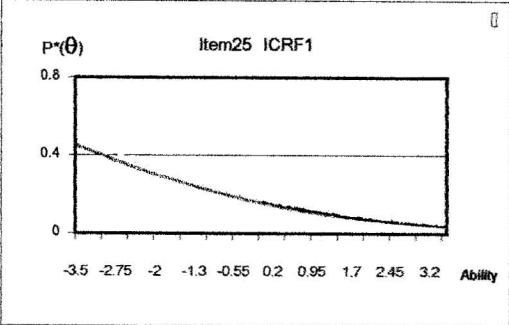
ข้อ 24 ลำดับขั้นที่ 5



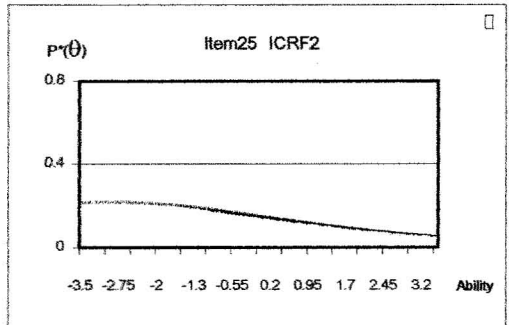
ข้อ 24 ลำดับขั้นที่ 6



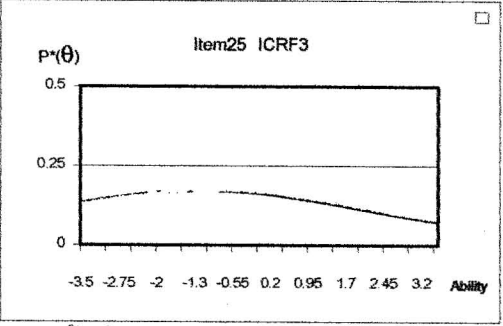
ข้อ 25 ลำดับขั้นที่ 1



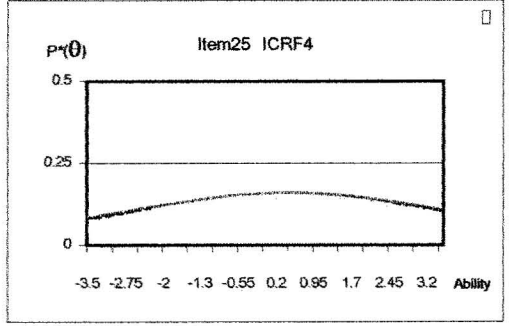
ข้อ 25 ลำดับขั้นที่ 2



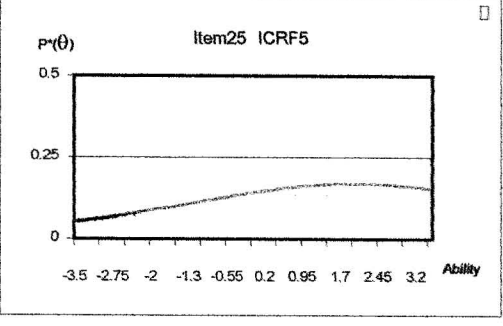
ข้อ 25 ลำดับขั้นที่ 3



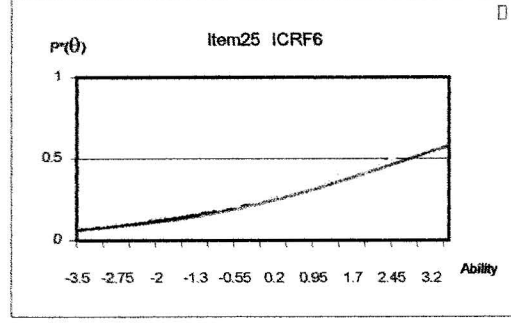
ข้อ 25 ลำดับขั้นที่ 4



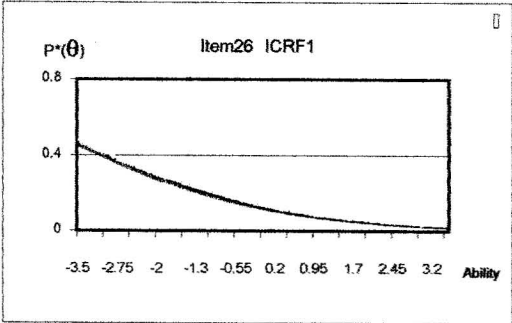
ข้อ 25 ลำดับขั้นที่ 5



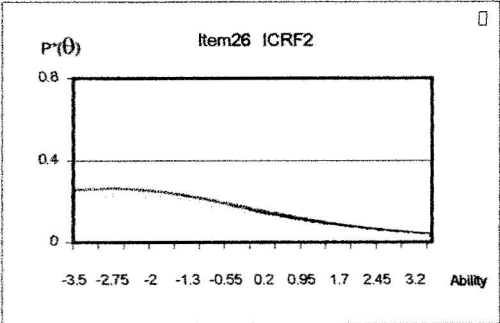
ข้อ 25 ลำดับขั้นที่ 6



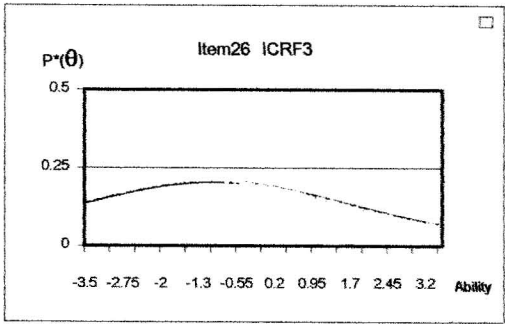
ข้อ 26 ลำดับขั้นที่ 1



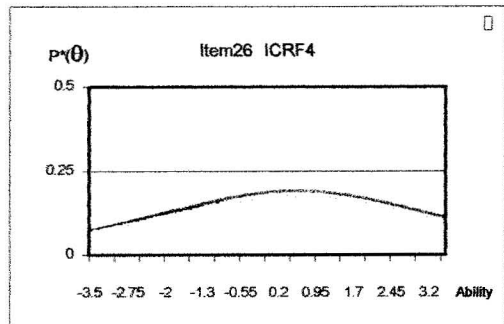
ลำดับขั้นที่ 2



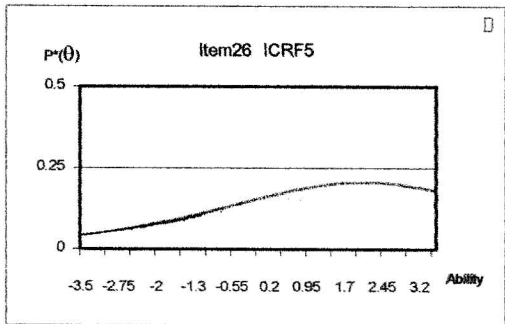
ลำดับขั้นที่ 3



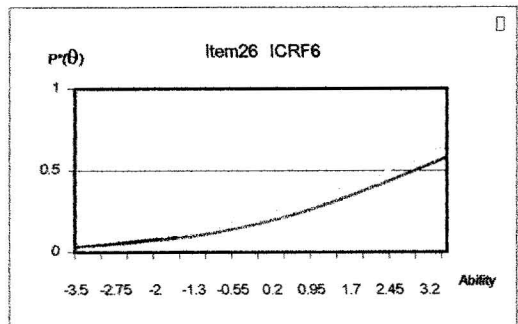
ลำดับขั้นที่ 4



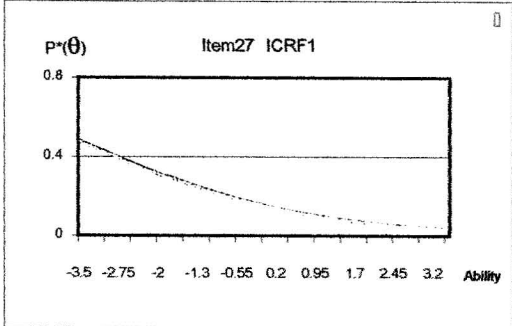
ลำดับขั้นที่ 5



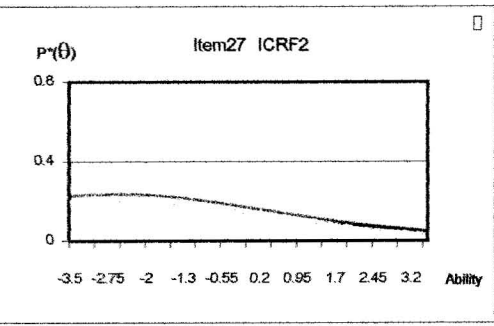
ลำดับขั้นที่ 6



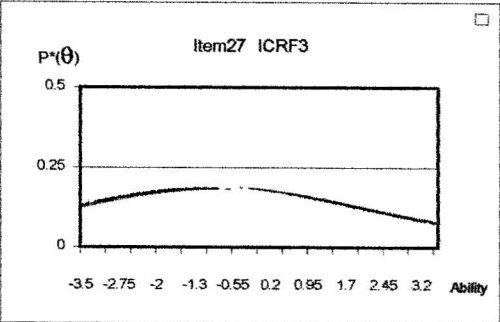
ข้อ 27 ลำดับขั้นที่ 1



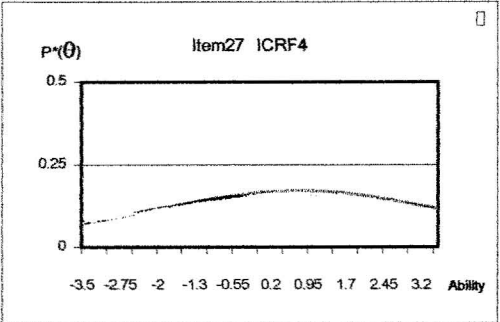
ลำดับขั้นที่ 2



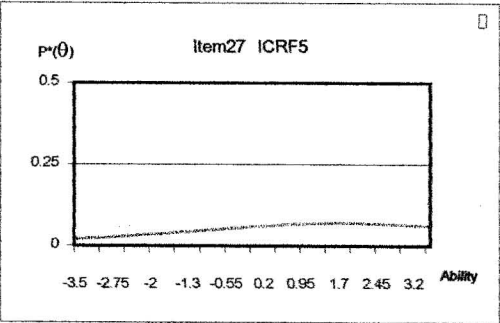
ลำดับชั้นที่ 3



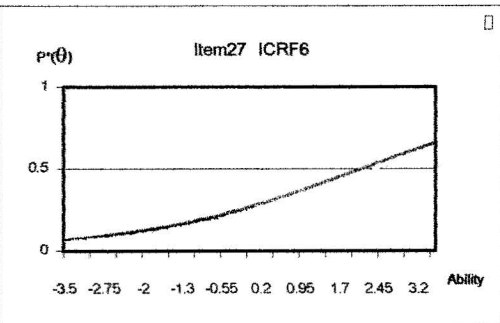
ลำดับชั้นที่ 4



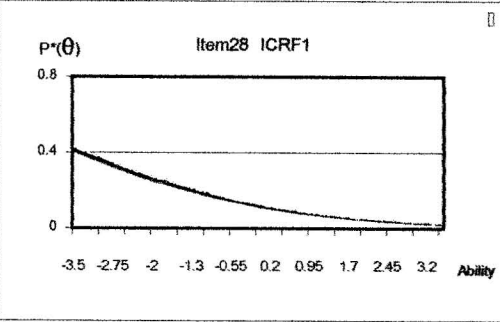
ลำดับชั้นที่ 5



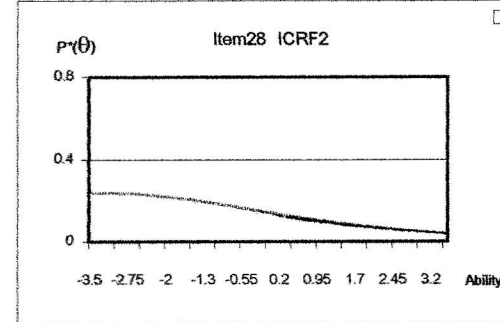
ลำดับชั้นที่ 6



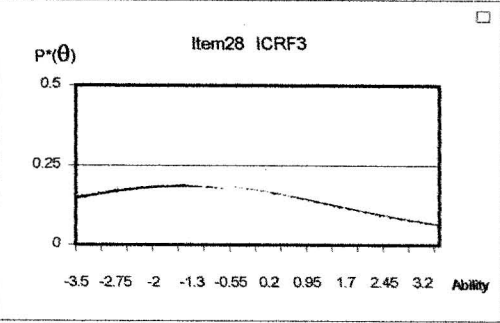
ข้อ 28 ลำดับชั้นที่ 1



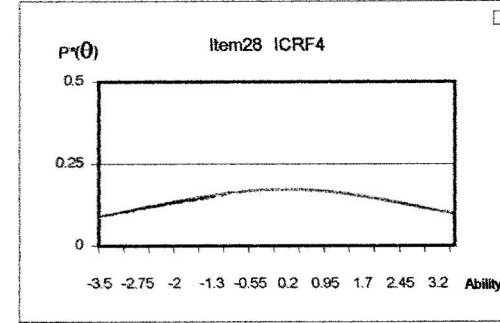
ลำดับชั้นที่ 2



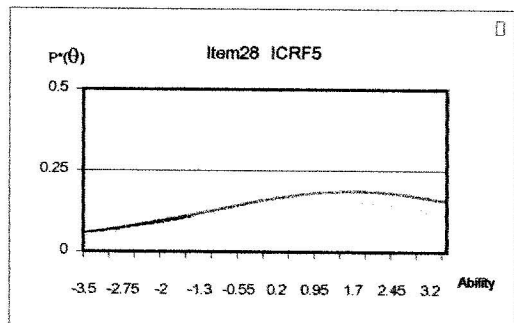
ลำดับชั้นที่ 3



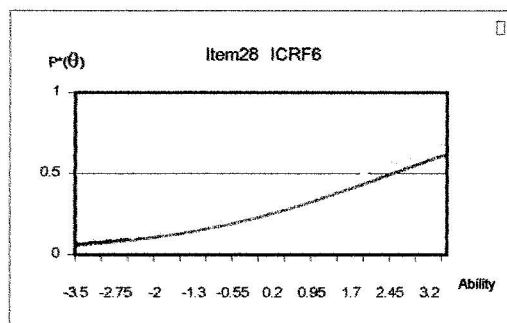
ลำดับชั้นที่ 4



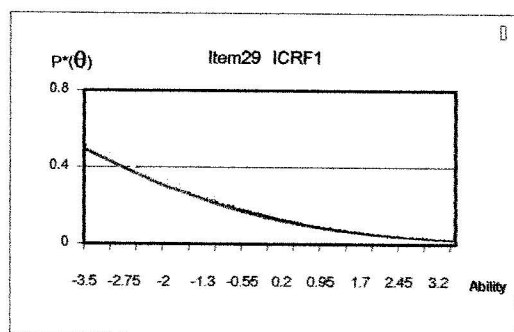
ลำดับขั้นที่ 5



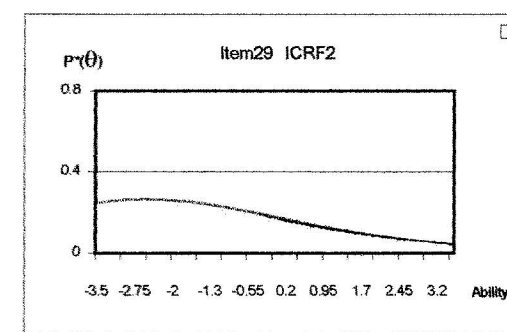
ลำดับขั้นที่ 6



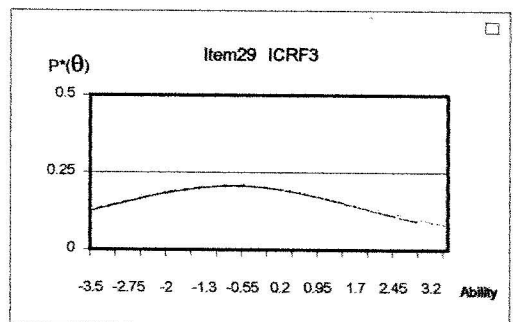
ข้อ 29 ลำดับขั้นที่ 1



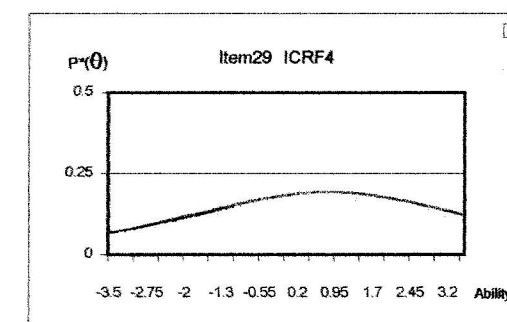
ลำดับขั้นที่ 2



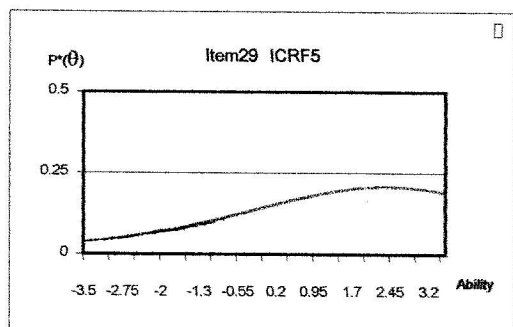
ลำดับขั้นที่ 3



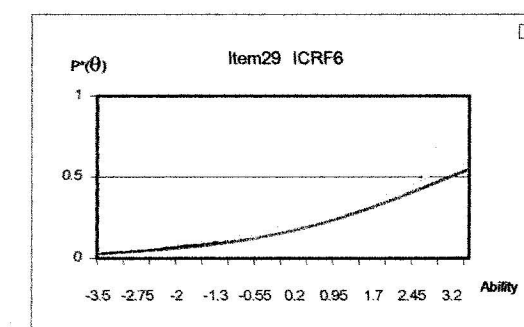
ลำดับขั้นที่ 4



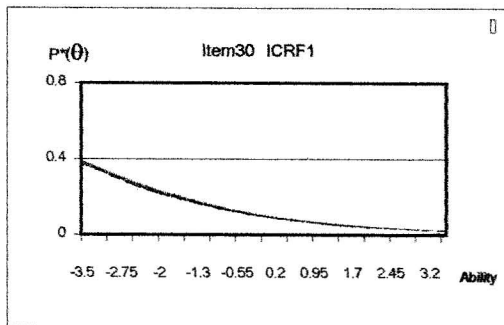
ลำดับขั้นที่ 5



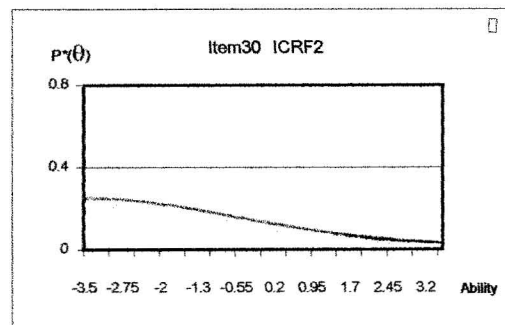
ลำดับขั้นที่ 6



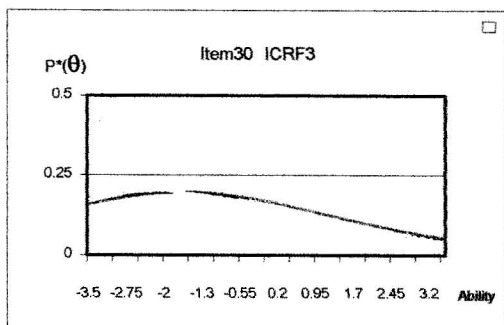
ข้อ 30 ลำดับขั้นที่ 1



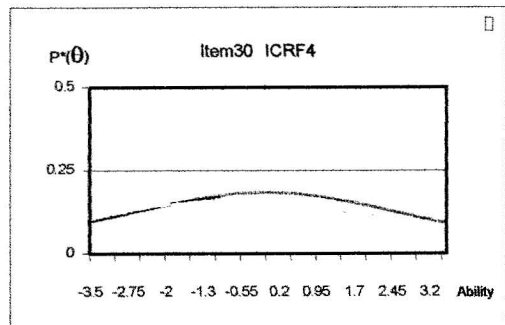
ลำดับขั้นที่ 2



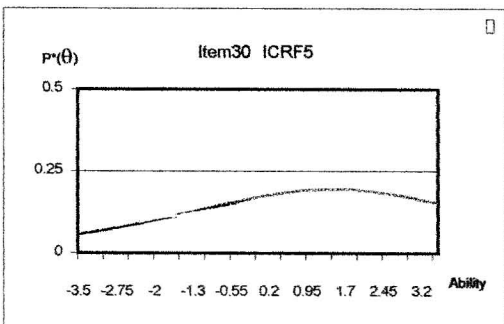
ลำดับขั้นที่ 3



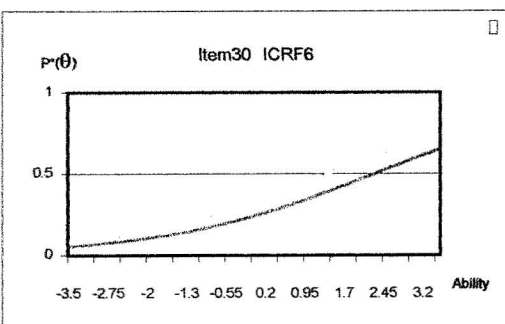
ลำดับขั้นที่ 4



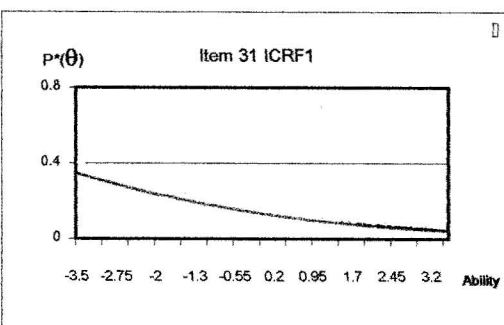
ลำดับขั้นที่ 5



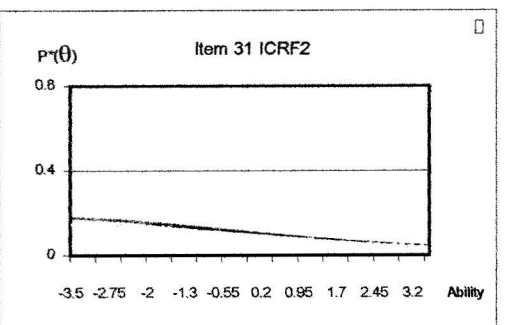
ลำดับขั้นที่ 6



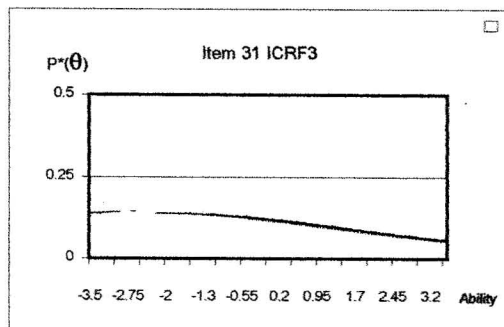
ข้อ 31 ลำดับขั้นที่ 1



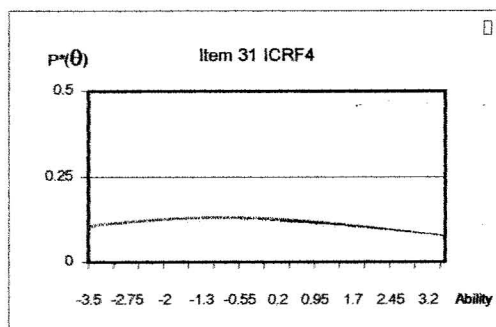
ลำดับขั้นที่ 2



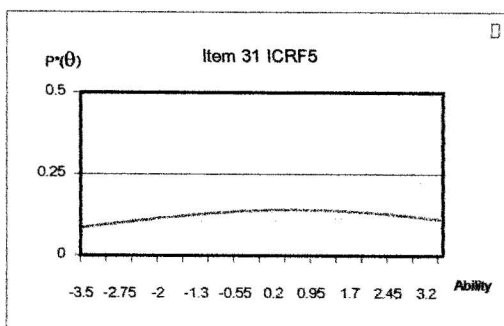
ลำดับขั้นที่ 3



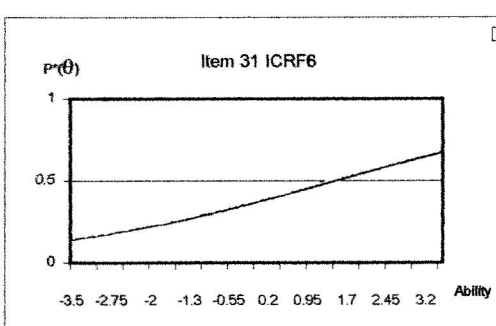
ลำดับขั้นที่ 4



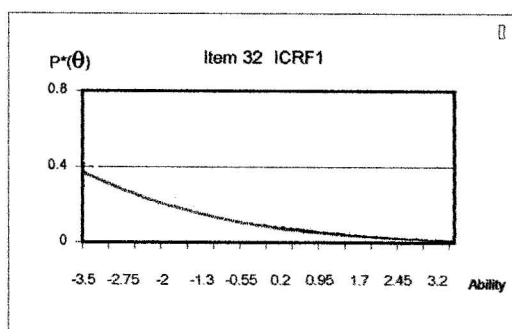
ลำดับขั้นที่ 5



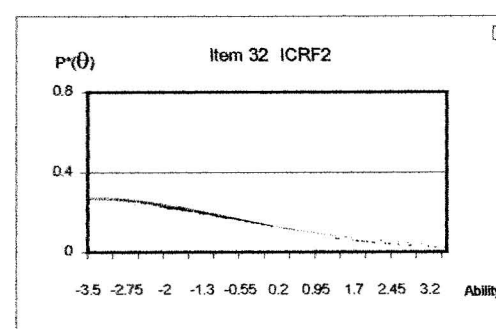
ลำดับขั้นที่ 6



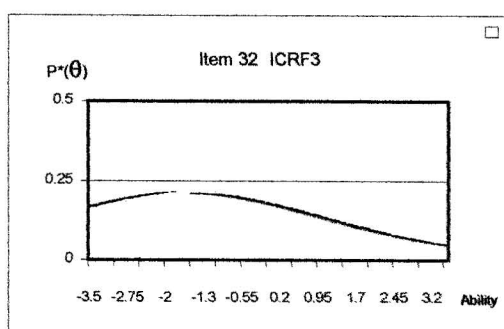
ข้อ 32 ลำดับขั้นที่ 1



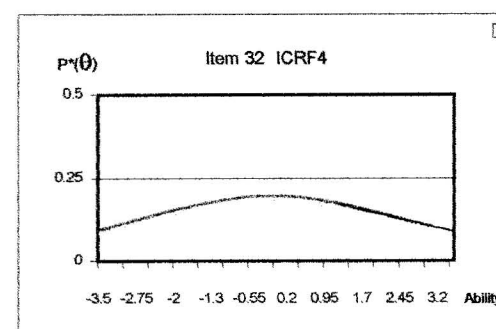
ลำดับขั้นที่ 2



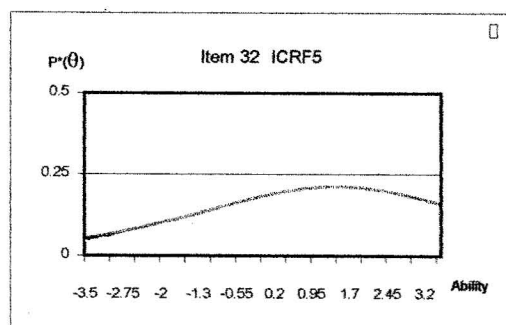
ลำดับขั้นที่ 3



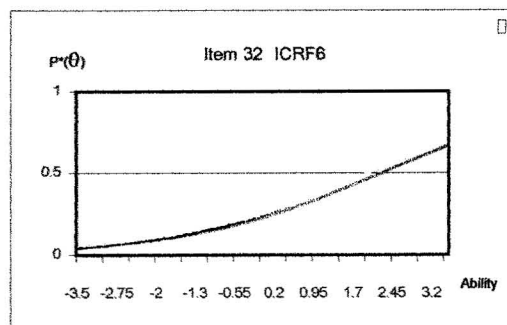
ลำดับขั้นที่ 4



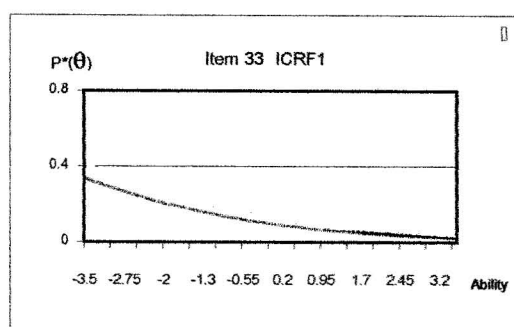
ลำดับขั้นที่ 5



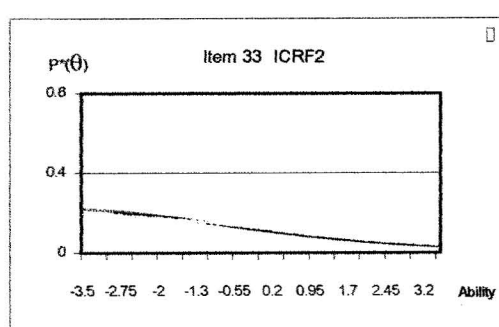
ลำดับขั้นที่ 6



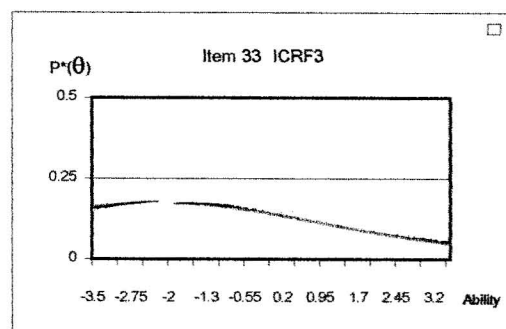
ข้อ 33 ลำดับขั้นที่ 1



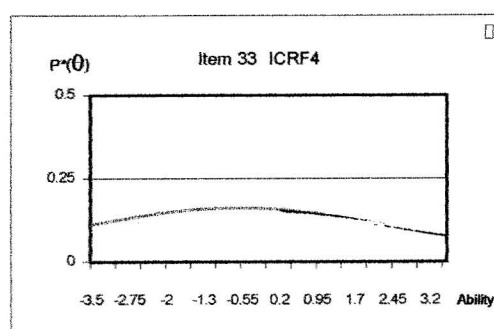
ลำดับขั้นที่ 2



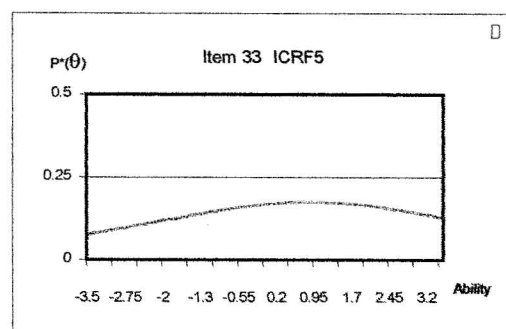
ลำดับขั้นที่ 3



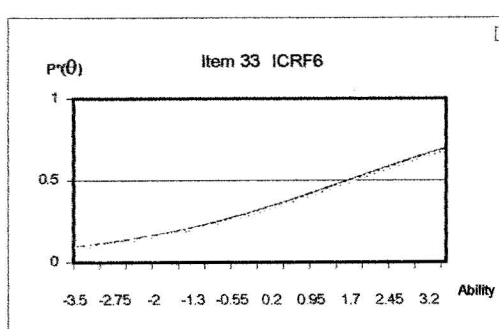
ลำดับขั้นที่ 4



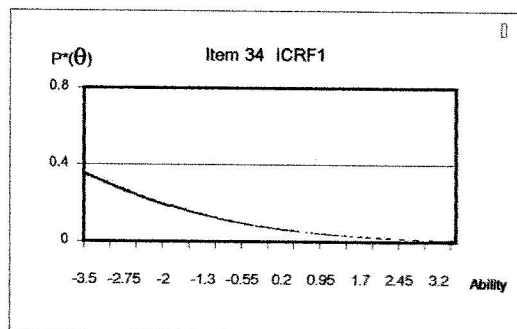
ลำดับขั้นที่ 5



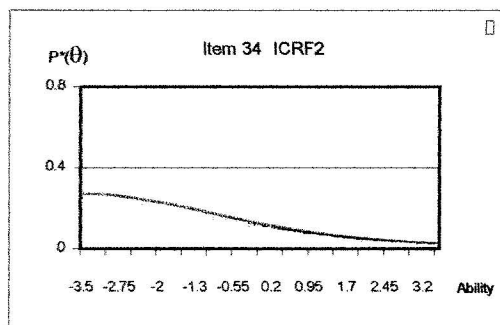
ลำดับขั้นที่ 6



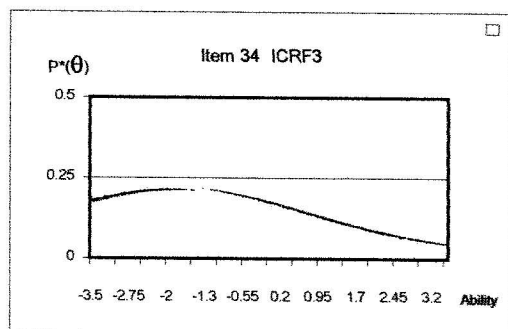
ข้อ 34 ลำดับขั้นที่ 1



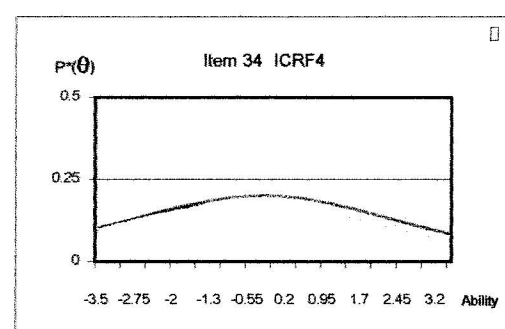
ลำดับขั้นที่ 2



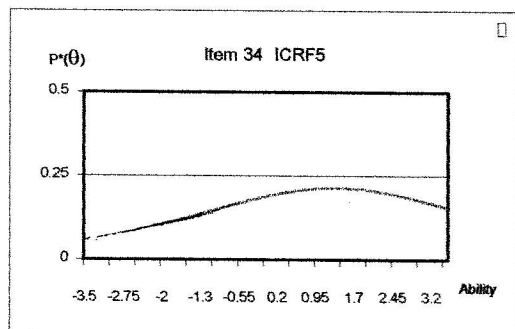
ลำดับขั้นที่ 3



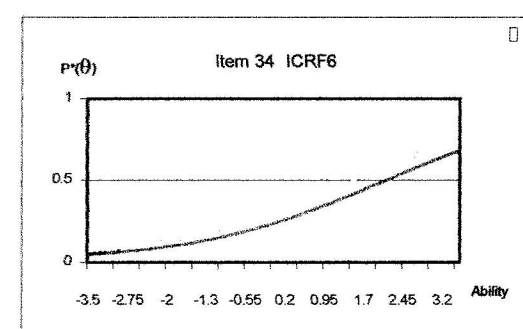
ลำดับขั้นที่ 4



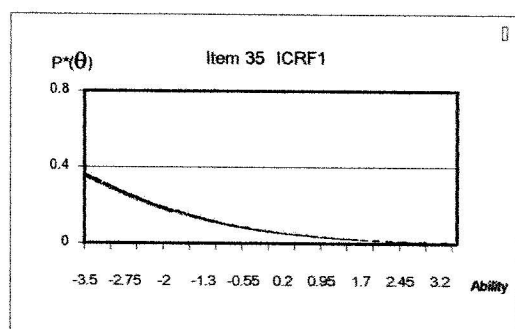
ลำดับขั้นที่ 5



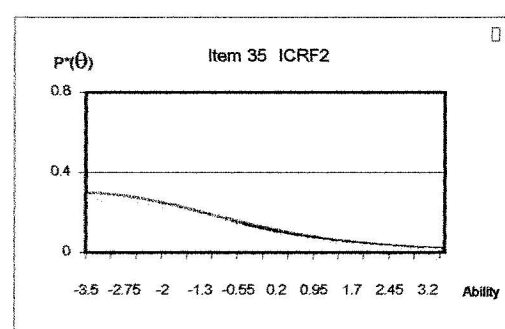
ลำดับขั้นที่ 6



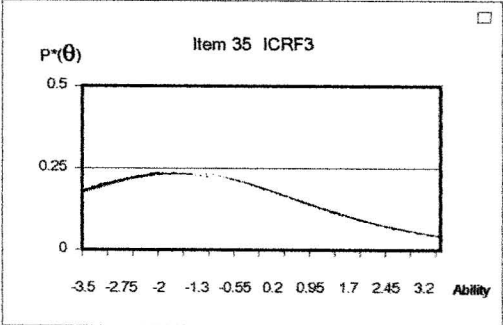
ข้อ 35 ลำดับขั้นที่ 1



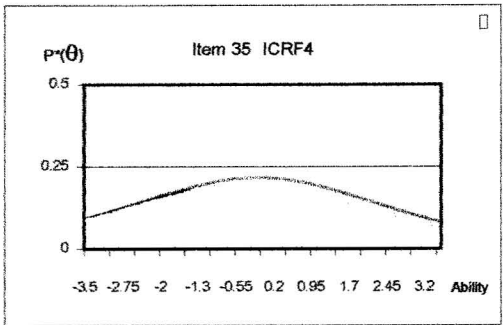
ลำดับขั้นที่ 2



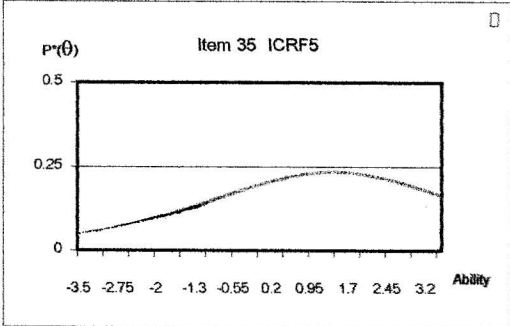
ลำดับชั้นที่ 3



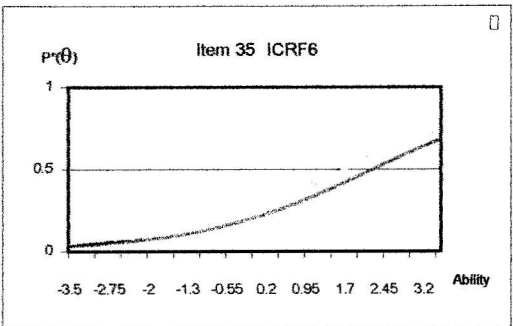
ลำดับชั้นที่ 4



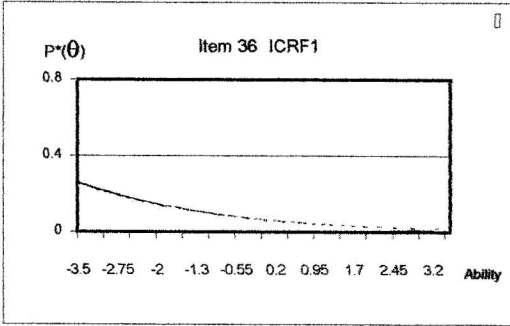
ลำดับชั้นที่ 5



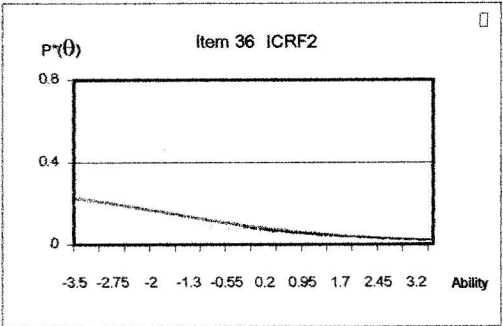
ลำดับชั้นที่ 6



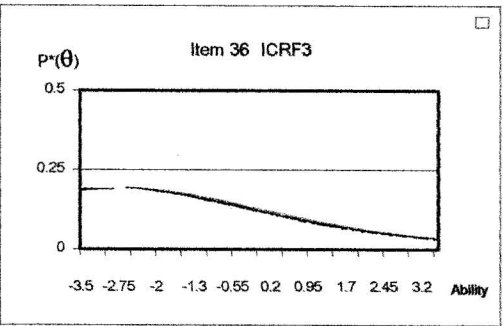
ข้อ 36 ลำดับชั้นที่ 1



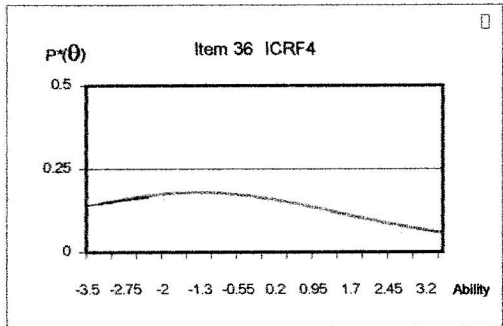
ลำดับชั้นที่ 2



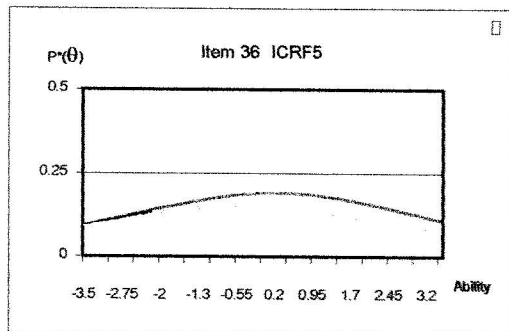
ลำดับชั้นที่ 3



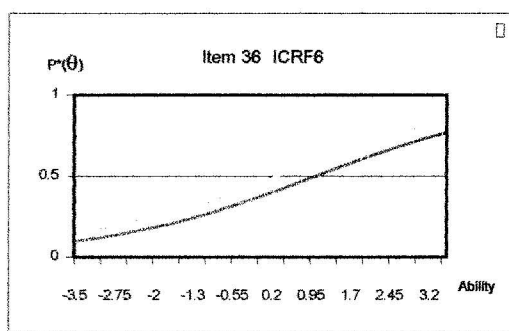
ลำดับชั้นที่ 4



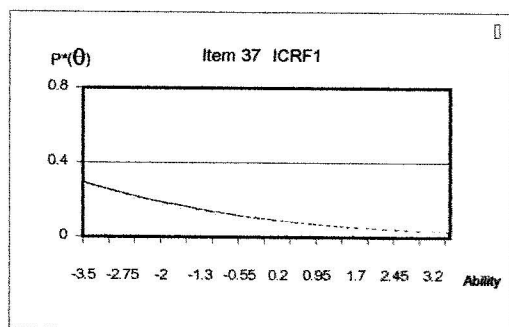
ลำดับชั้นที่ 5



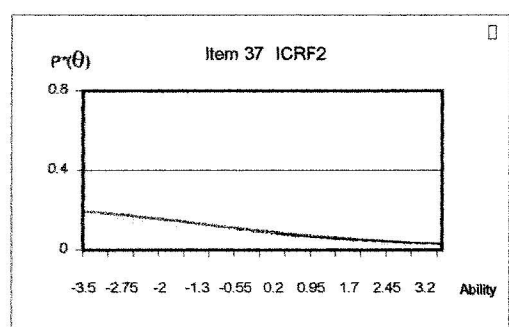
ลำดับชั้นที่ 6



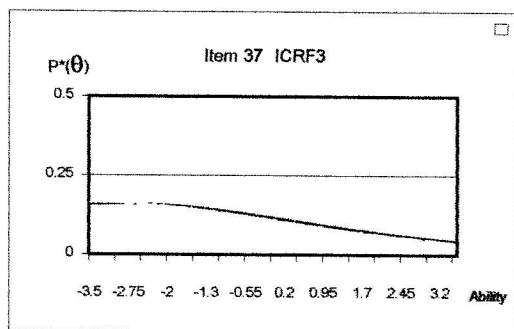
ข้อ 37 ลำดับชั้นที่ 1



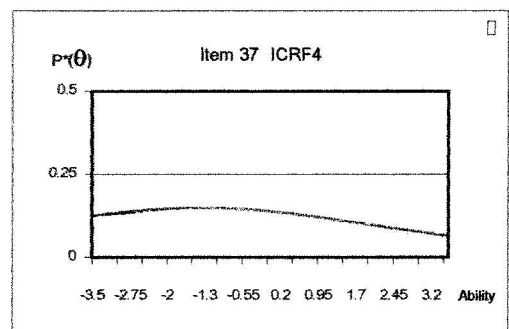
ลำดับชั้นที่ 2



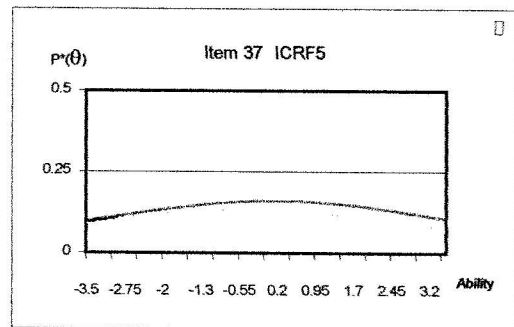
ลำดับชั้นที่ 3



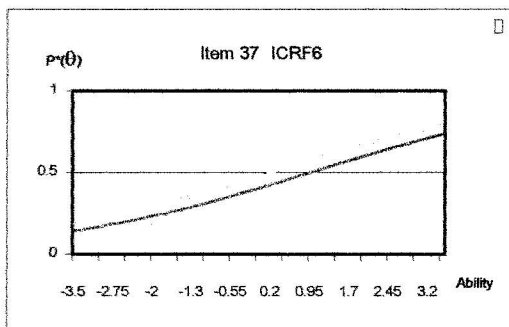
ลำดับชั้นที่ 4



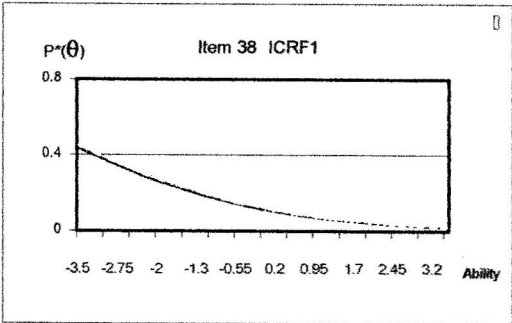
ลำดับชั้นที่ 5



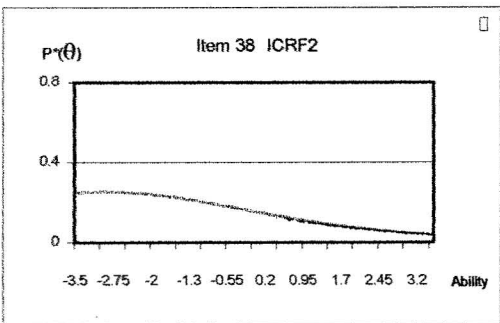
ลำดับชั้นที่ 6



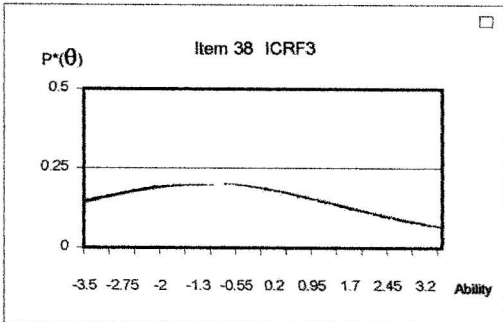
ข้อ 38 ลำดับขั้นที่ 1



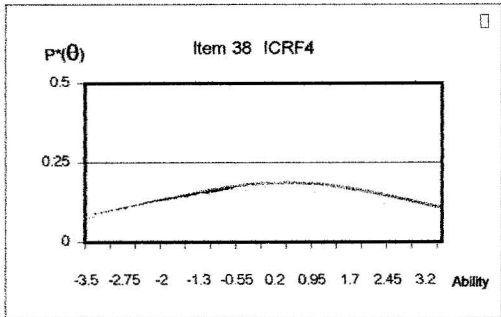
ลำดับขั้นที่ 2



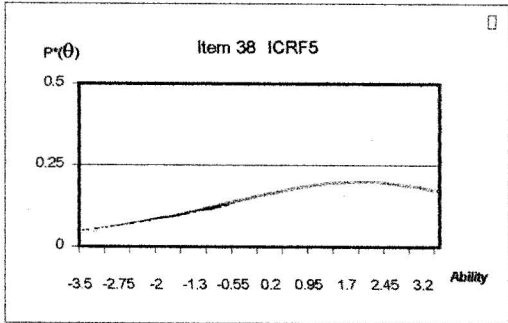
ลำดับขั้นที่ 3



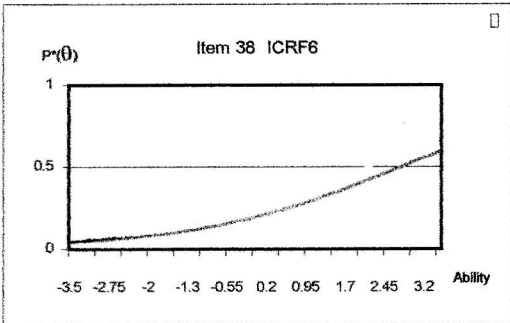
ลำดับขั้นที่ 4



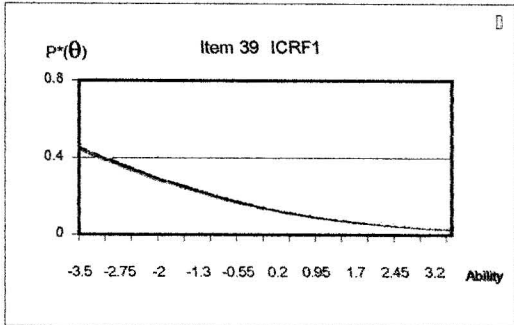
ลำดับขั้นที่ 5



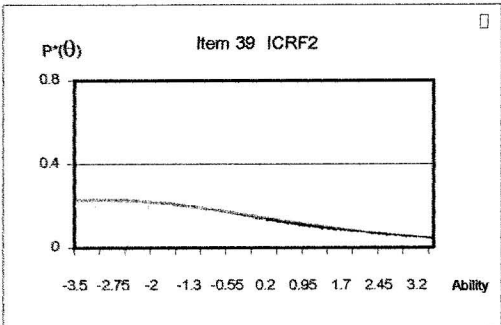
ลำดับขั้นที่ 6



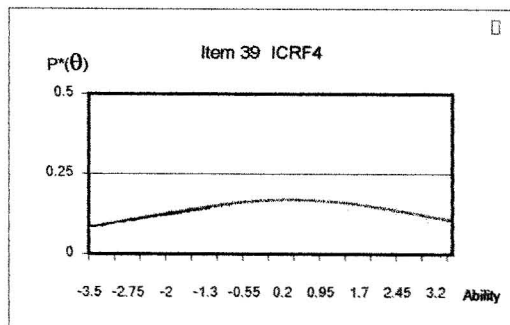
ข้อ 39 ลำดับขั้นที่ 1



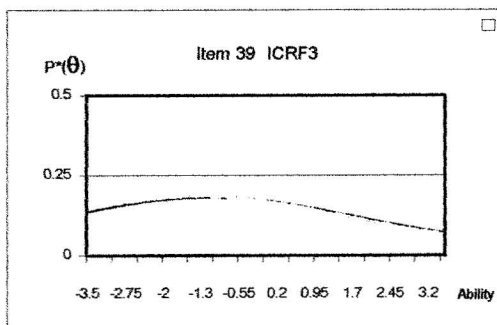
ลำดับขั้นที่ 2



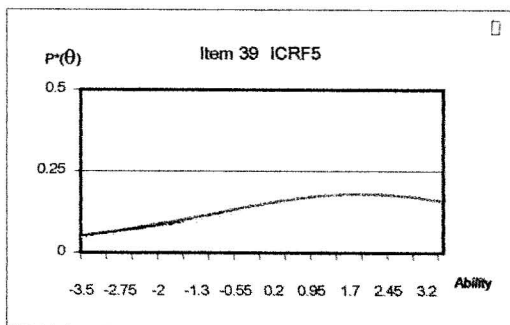
ลำดับชั้นที่ 3



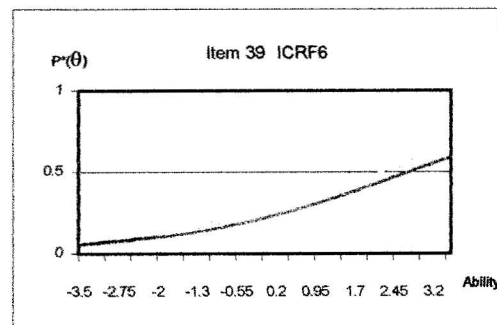
ลำดับชั้นที่ 4



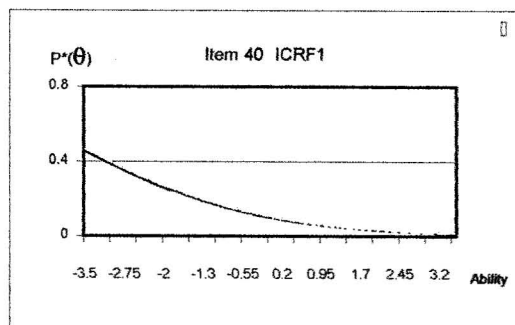
ลำดับชั้นที่ 5



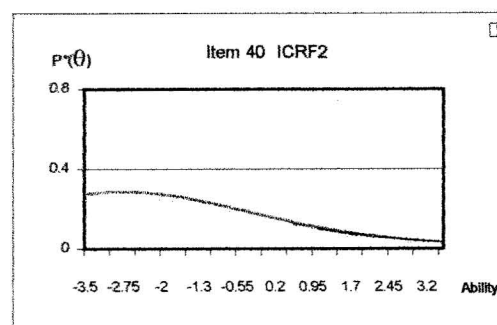
ลำดับชั้นที่ 6



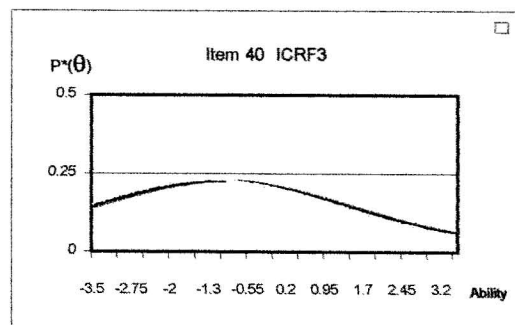
ข้อ 40 ลำดับชั้นที่ 1



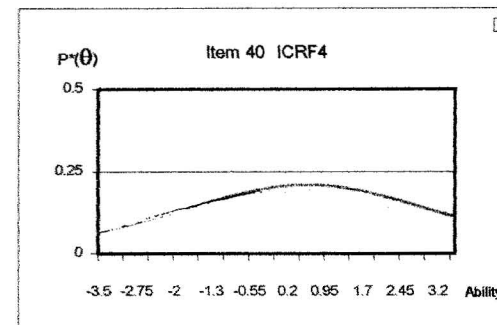
ลำดับชั้นที่ 2



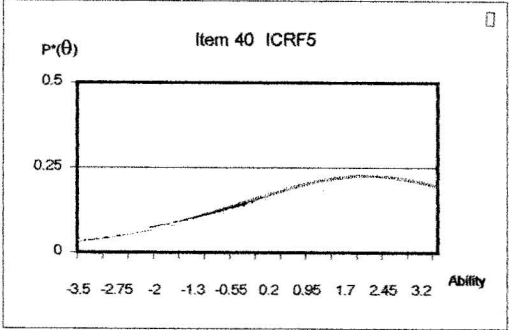
ลำดับชั้นที่ 3



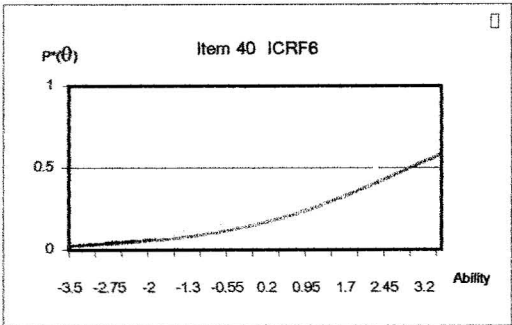
ลำดับชั้นที่ 4



ลำดับชั้นที่ 5

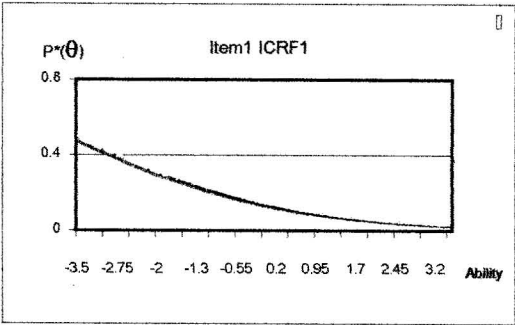


ลำดับชั้นที่ 6

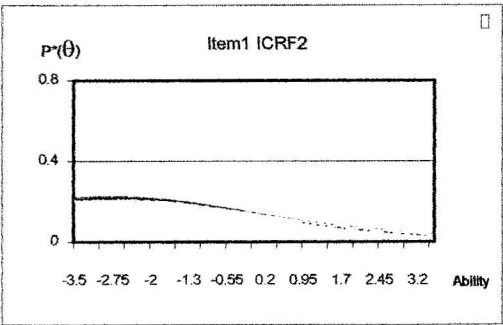


ฟังก์ชันผลการตอบลำดับชั้นข้อสอบ (ICRF) ที่วิเคราะห์ด้วยวิธี GPCM ในแบบทดสอบขนาด 30 ข้อ
เมื่อกลุ่มตัวอย่างเพศชายต่อเพศหญิง เท่ากับ 1,000 คน ต่อ 1,000 คน

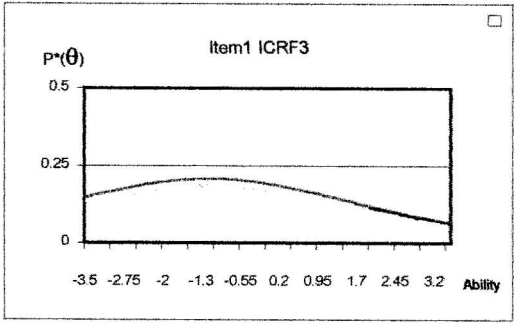
ข้อ 1 ลำดับชั้นที่ 1



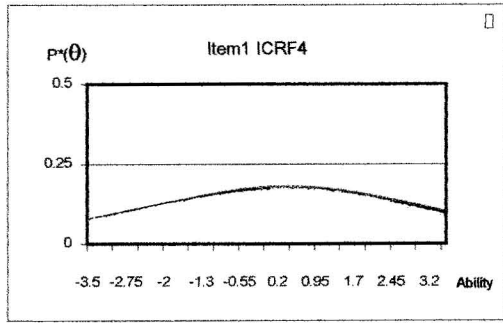
ลำดับชั้นที่ 2



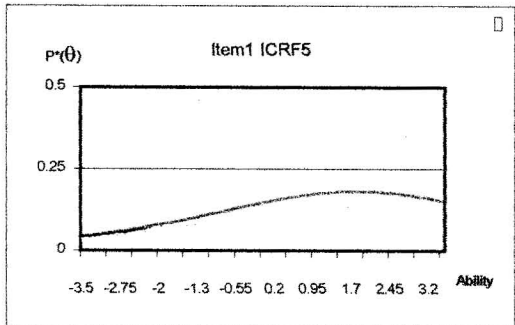
ลำดับชั้นที่ 3



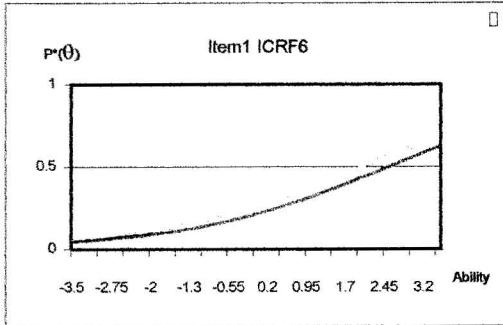
ลำดับชั้นที่ 4



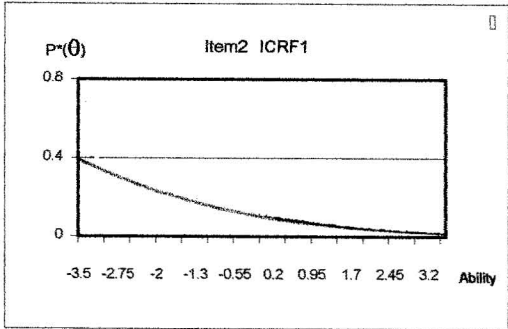
ลำดับชั้นที่ 5



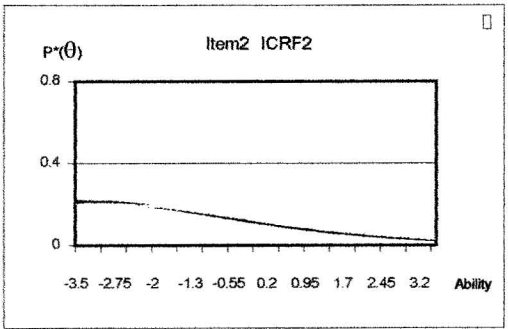
ลำดับชั้นที่ 6



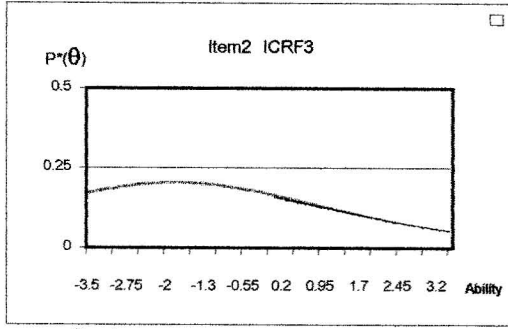
ข้อ 2 ลำดับขั้นที่ 1



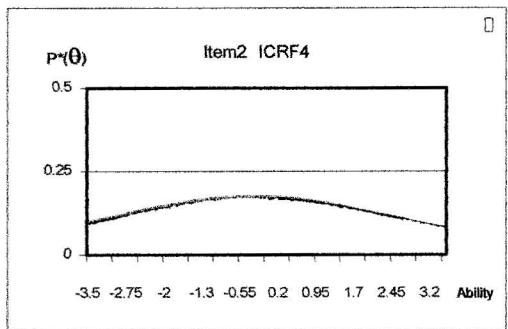
ลำดับขั้นที่ 2



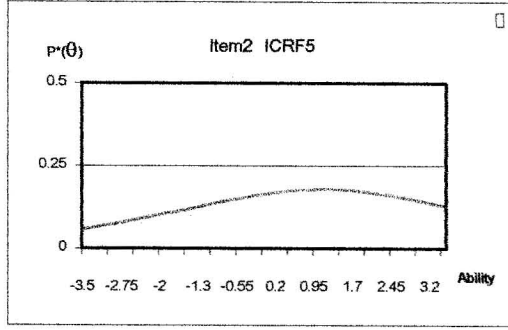
ลำดับขั้นที่ 3



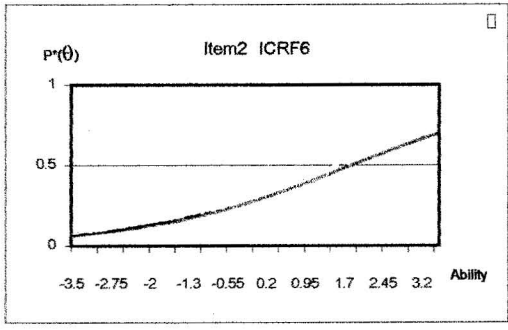
ลำดับขั้นที่ 4



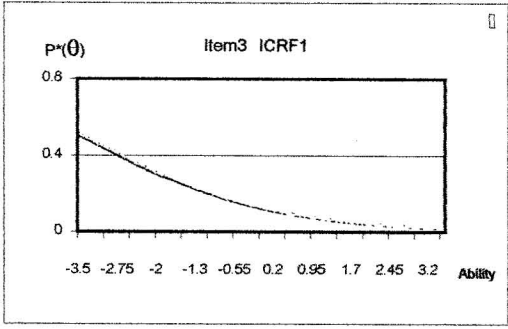
ลำดับขั้นที่ 5



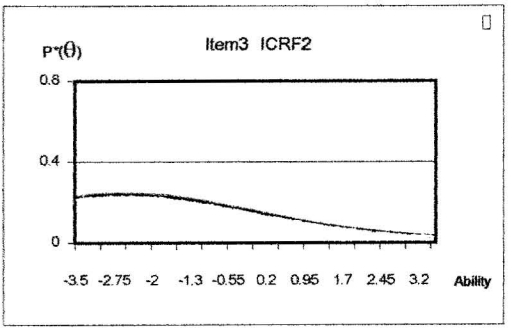
ลำดับขั้นที่ 6



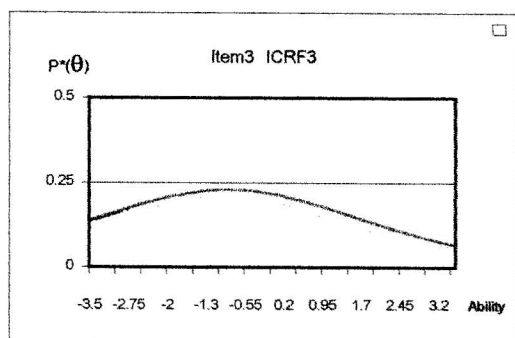
ข้อ 3 ลำดับขั้นที่ 1



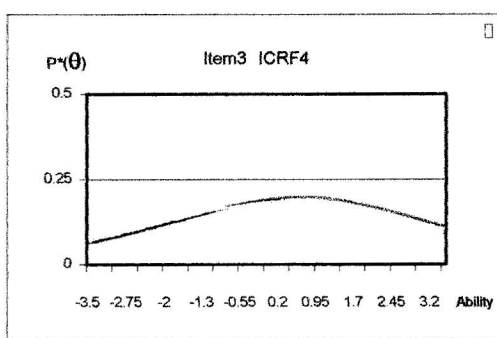
ลำดับขั้นที่ 2



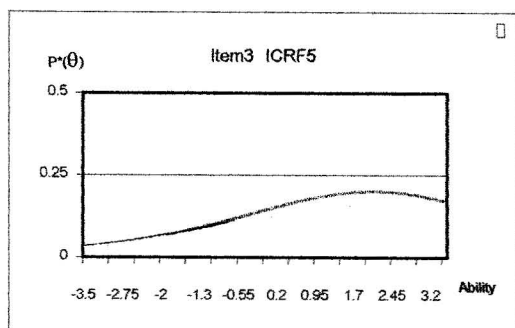
ลำดับชั้นที่ 3



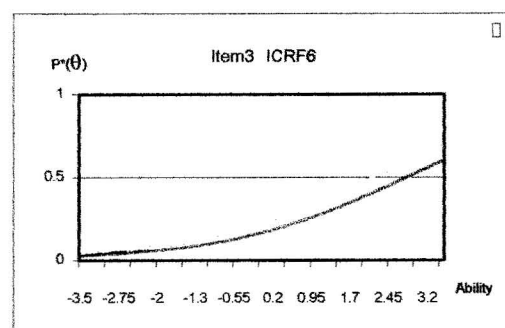
ลำดับชั้นที่ 4



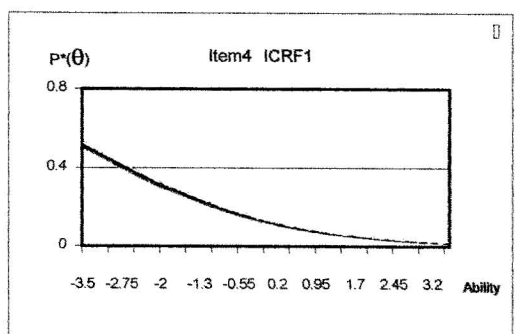
ลำดับชั้นที่ 5



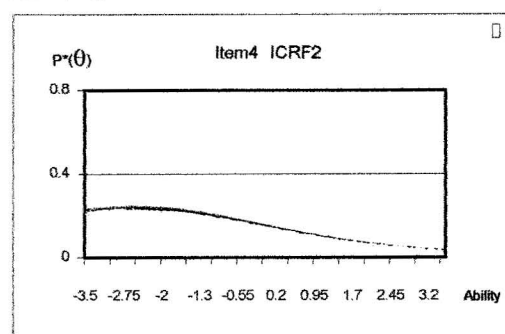
ลำดับชั้นที่ 6



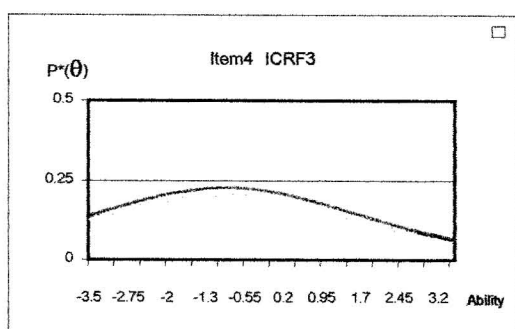
ข้อ 4 ลำดับชั้นที่ 1



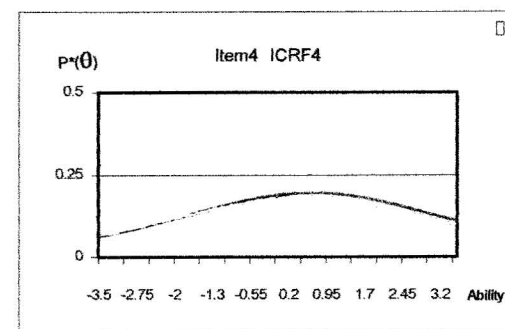
ลำดับชั้นที่ 2



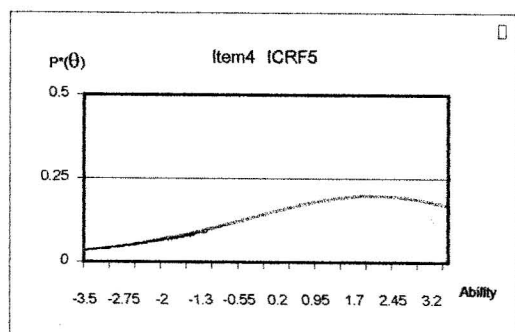
ลำดับชั้นที่ 3



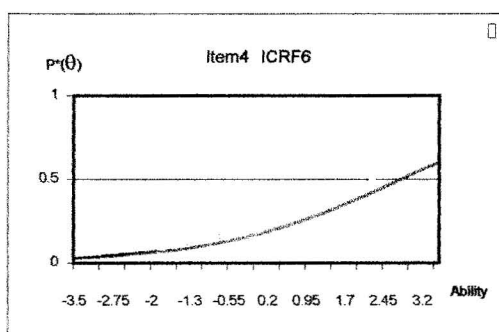
ลำดับชั้นที่ 4



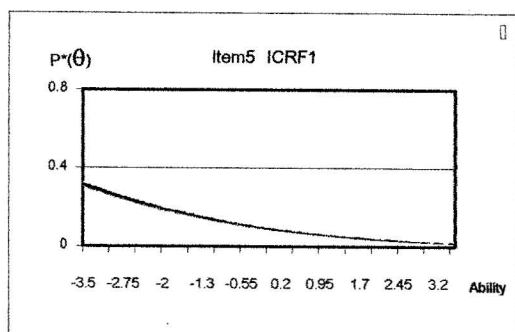
ลำดับขั้นที่ 5



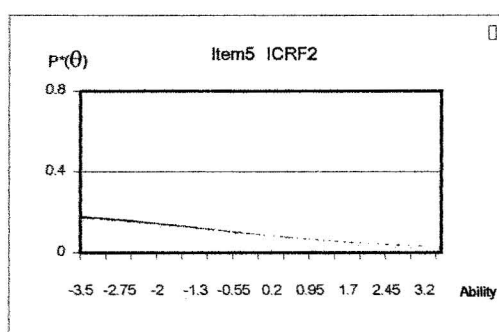
ลำดับขั้นที่ 6



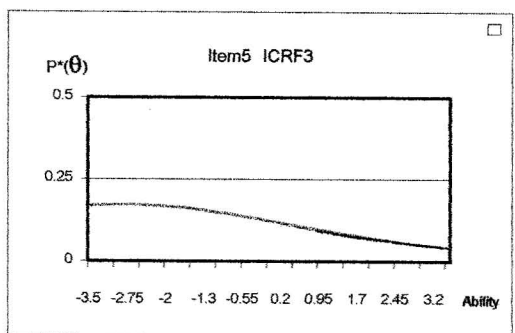
ข้อ 5 ลำดับขั้นที่ 1



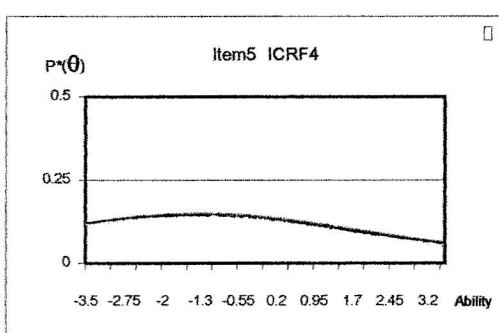
ลำดับขั้นที่ 2



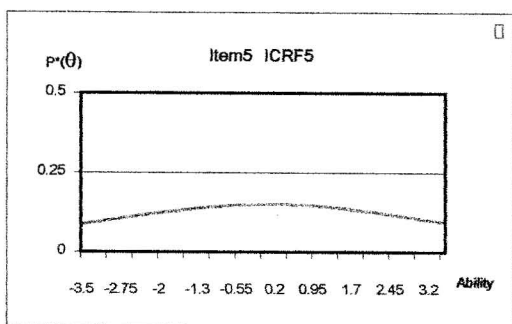
ลำดับขั้นที่ 3



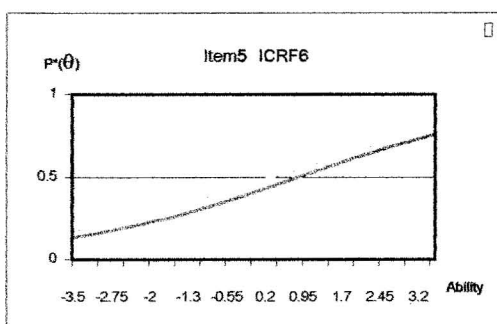
ลำดับขั้นที่ 4



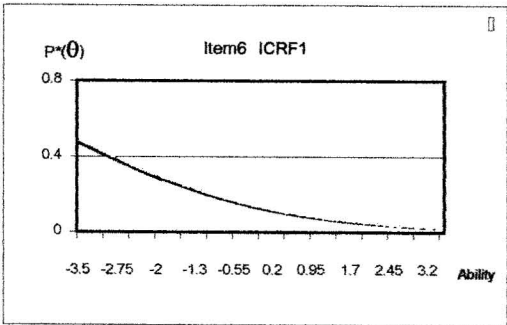
ลำดับขั้นที่ 5



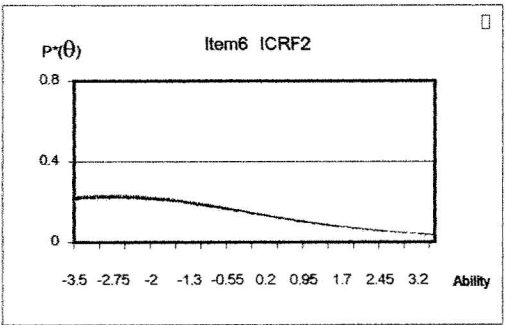
ลำดับขั้นที่ 6



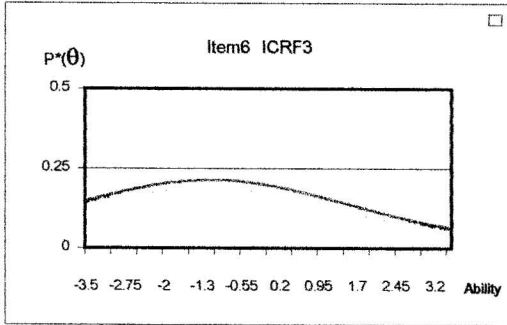
ข้อ 6 ลำดับขั้นที่ 1



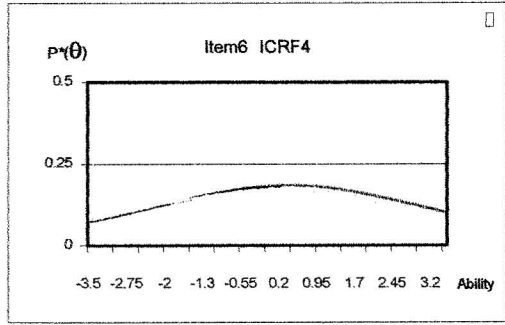
ลำดับขั้นที่ 2



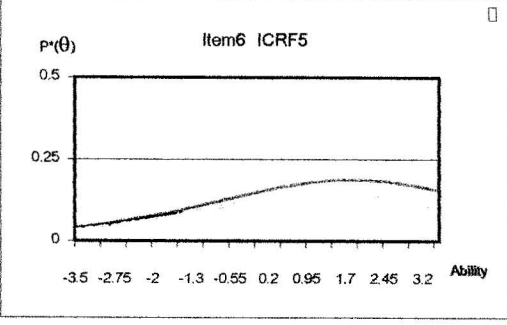
ลำดับขั้นที่ 3



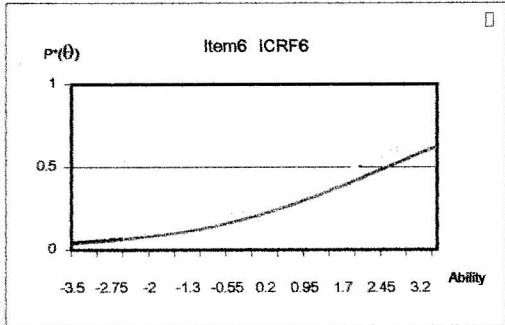
ลำดับขั้นที่ 4



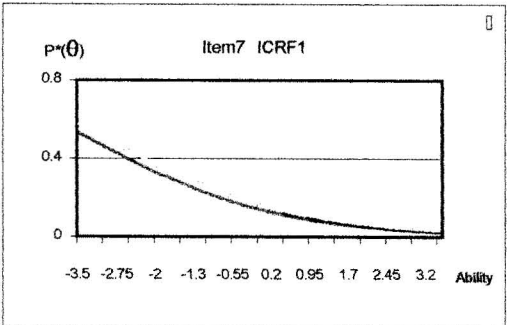
ลำดับขั้นที่ 5



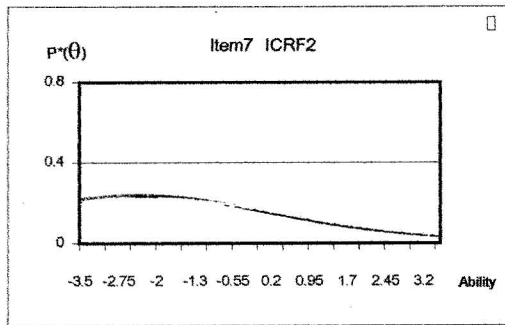
ลำดับขั้นที่ 6



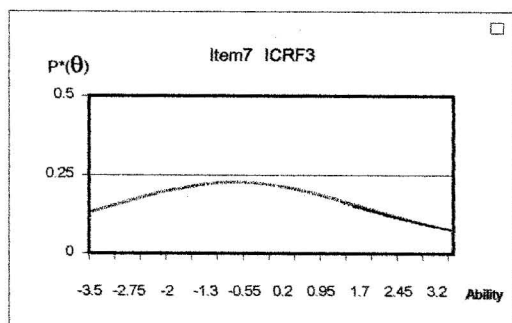
ข้อ 7 ลำดับขั้นที่ 1



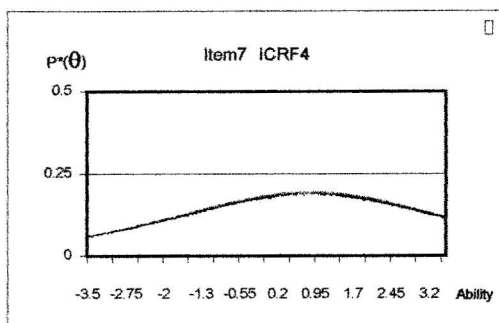
ลำดับขั้นที่ 2



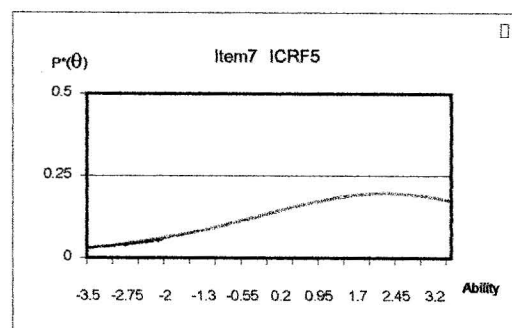
ลำดับชั้นที่ 3



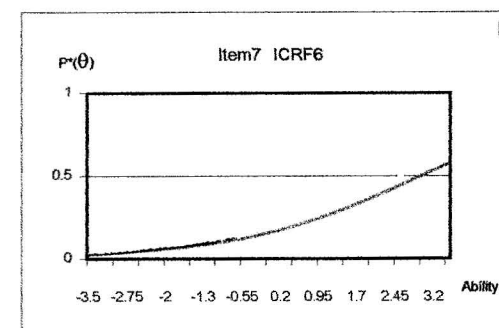
ลำดับชั้นที่ 4



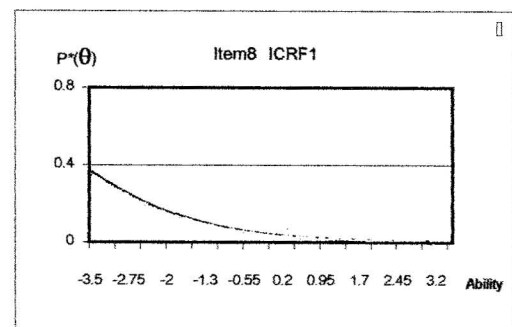
ลำดับชั้นที่ 5



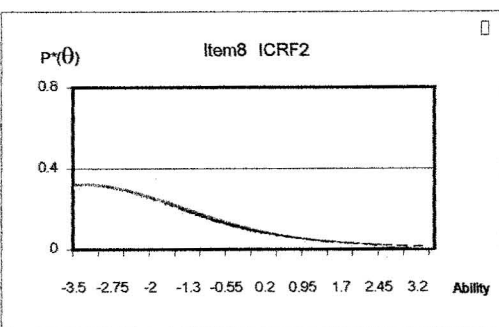
ลำดับชั้นที่ 6



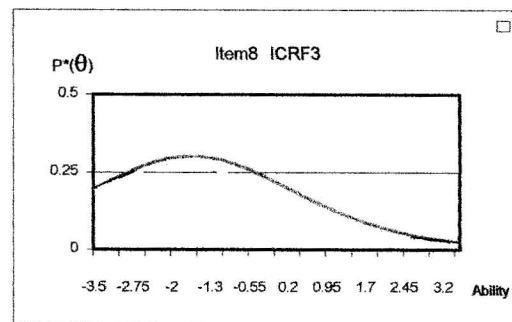
ข้อ 8 ลำดับชั้นที่ 1



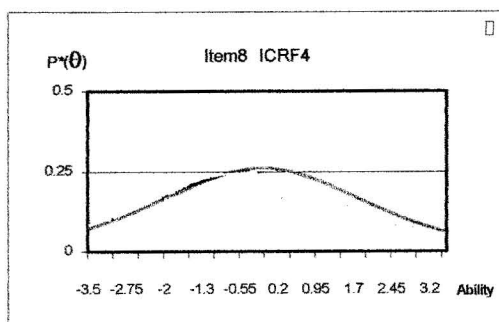
ลำดับชั้นที่ 2



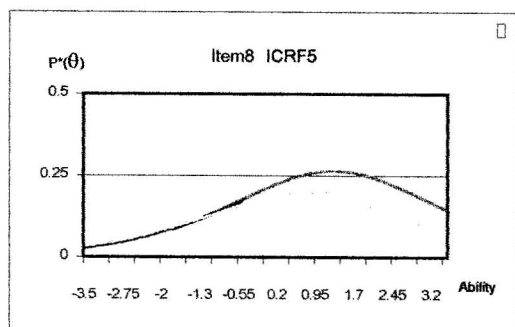
ลำดับชั้นที่ 3



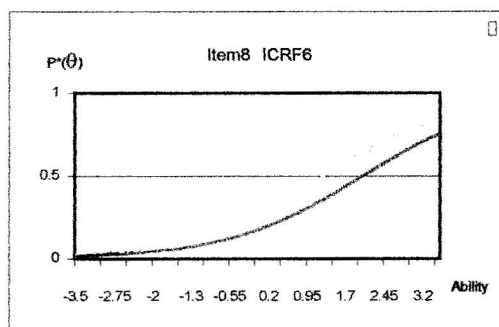
ลำดับชั้นที่ 4



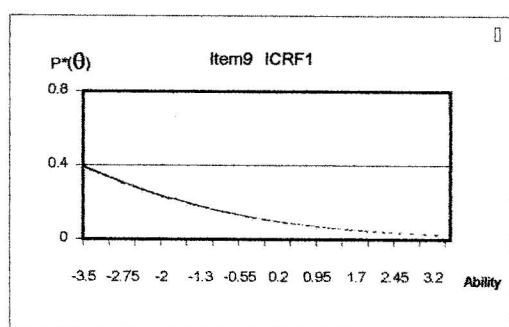
ลำดับขั้นที่ 5



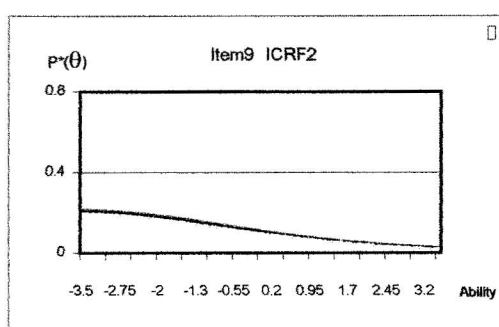
ลำดับขั้นที่ 6



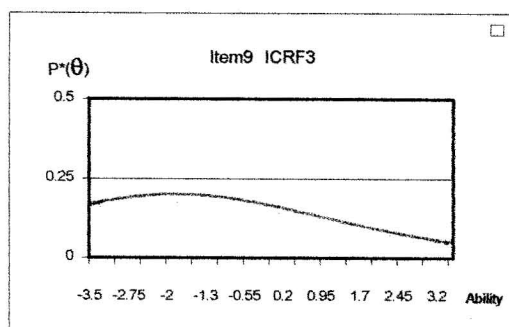
ข้อ 9 ลำดับขั้นที่ 1



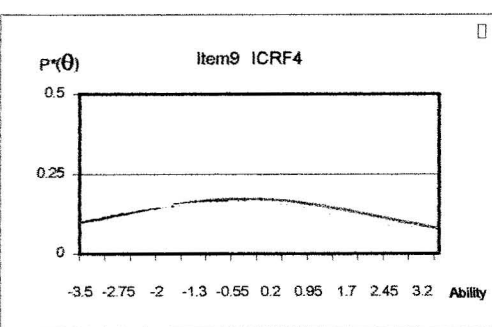
ลำดับขั้นที่ 2



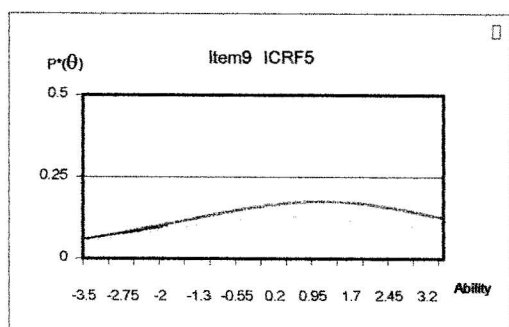
ลำดับขั้นที่ 3



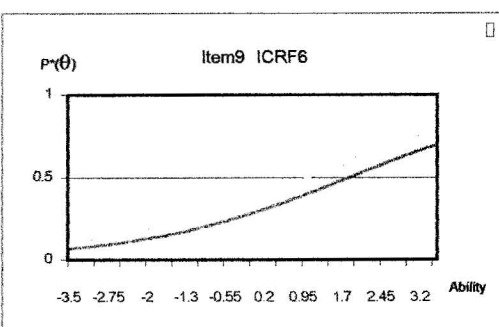
ลำดับขั้นที่ 4



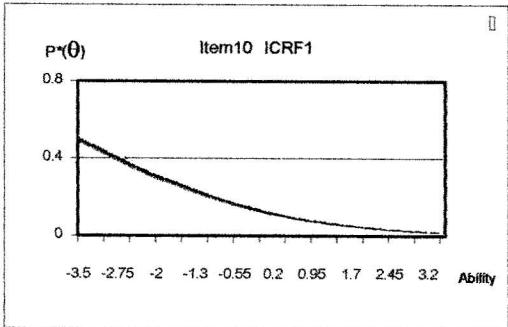
ลำดับขั้นที่ 5



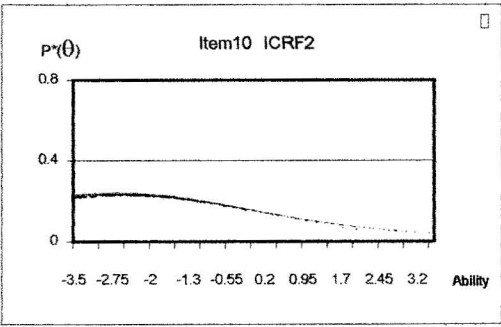
ลำดับขั้นที่ 6



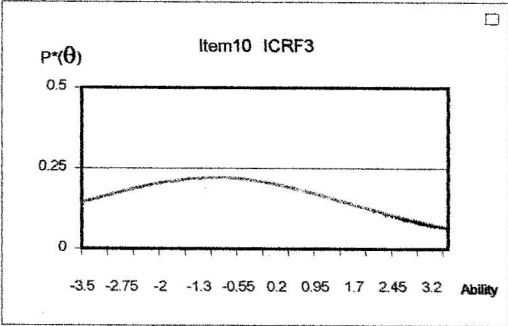
ข้อ 10 ลำดับขั้นที่ 1



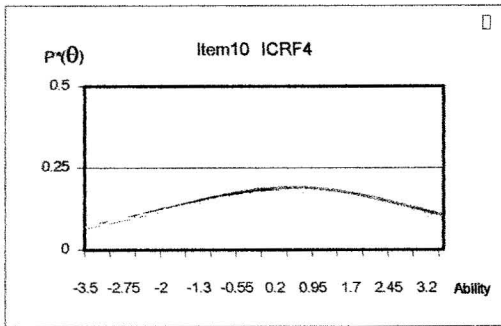
ลำดับขั้นที่ 2



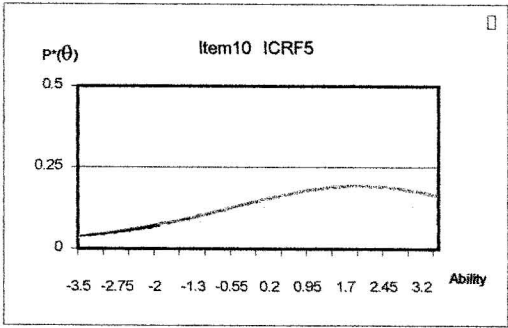
ลำดับขั้นที่ 3



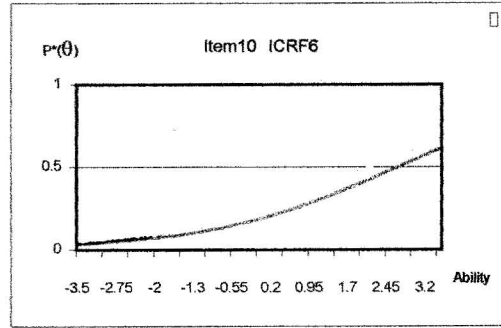
ลำดับขั้นที่ 4



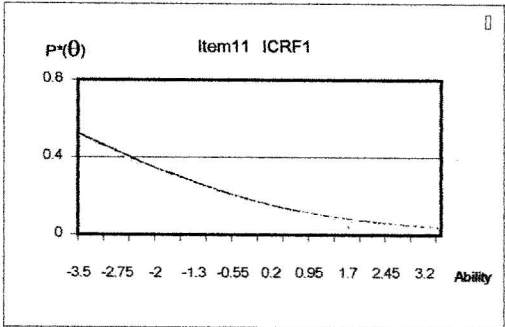
ลำดับขั้นที่ 5



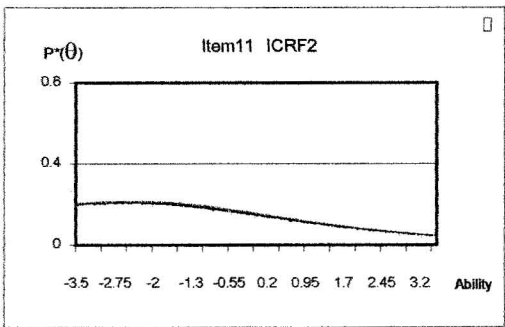
ลำดับขั้นที่ 6



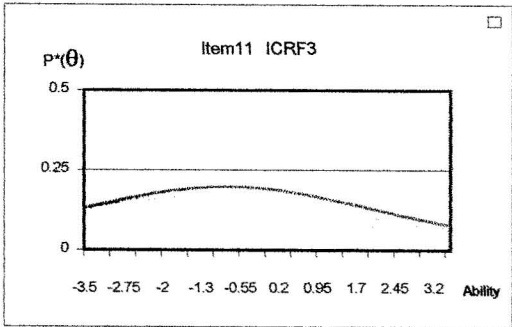
ข้อ 11 ลำดับขั้นที่ 1



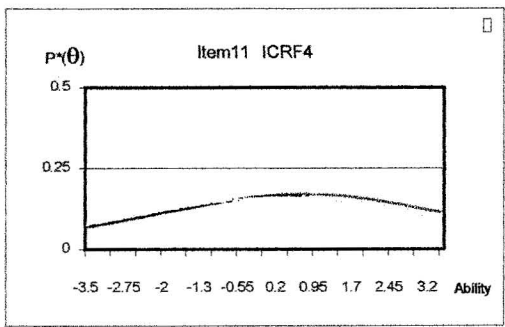
ลำดับขั้นที่ 2



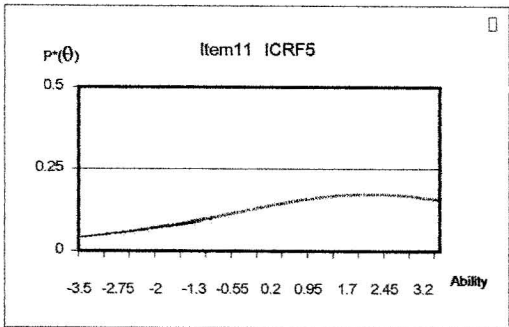
ลำดับขั้นที่ 3



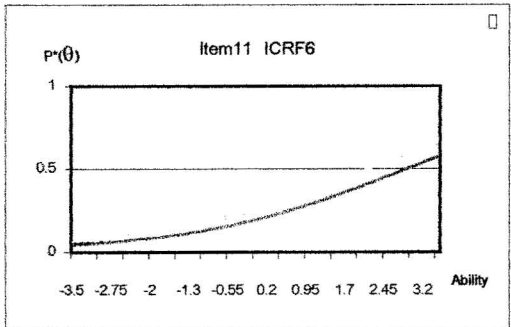
ลำดับขั้นที่ 4



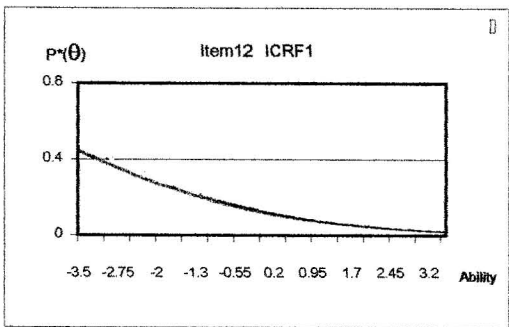
ลำดับขั้นที่ 5



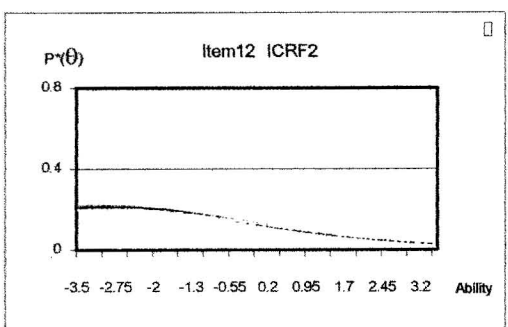
ลำดับขั้นที่ 6



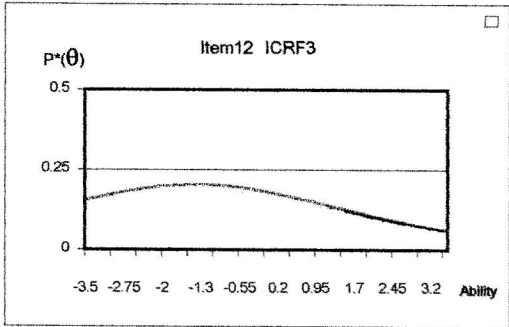
ข้อ 12 ลำดับขั้นที่ 1



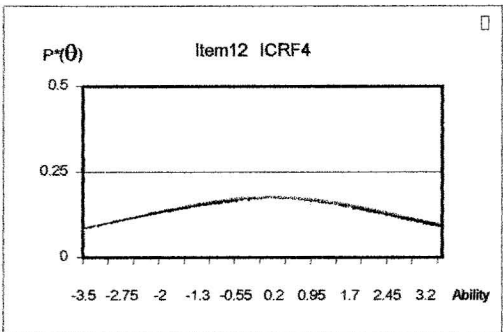
ลำดับขั้นที่ 2



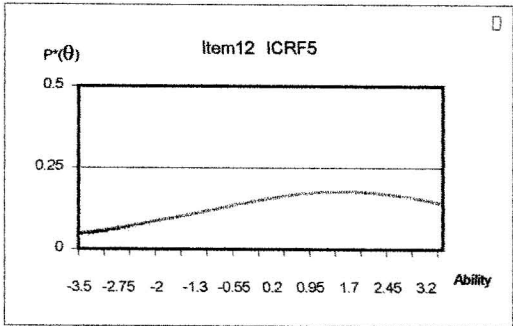
ลำดับชั้นที่ 3



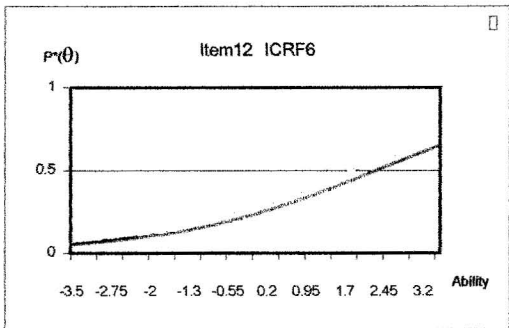
ลำดับชั้นที่ 4



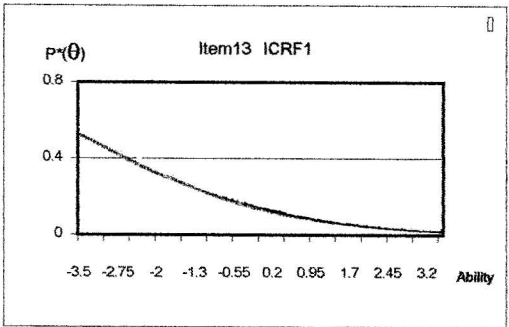
ลำดับชั้นที่ 5



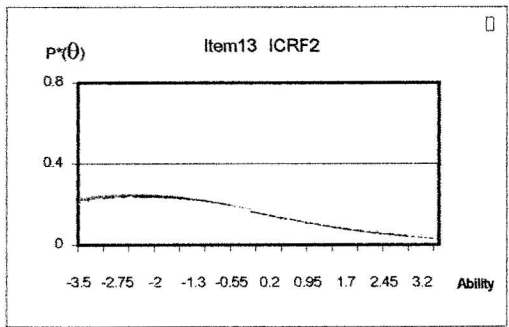
ลำดับชั้นที่ 6



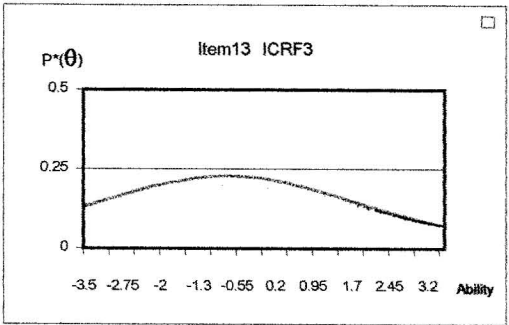
ข้อ 13 ลำดับชั้นที่ 1



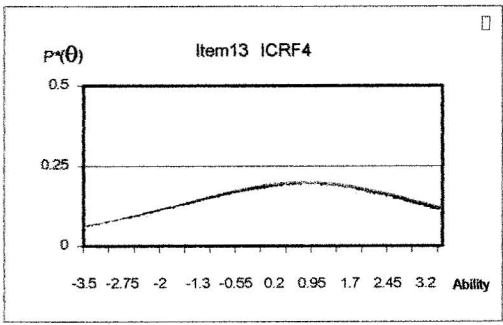
ลำดับชั้นที่ 2



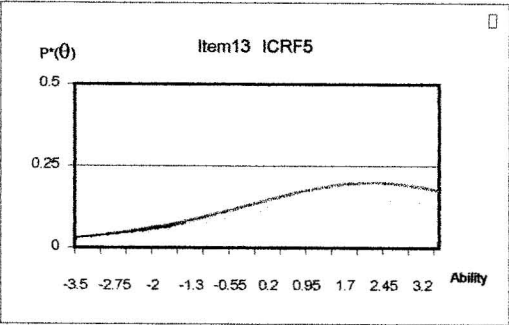
ลำดับชั้นที่ 3



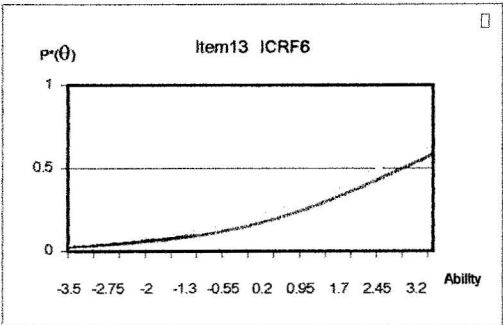
ลำดับชั้นที่ 4



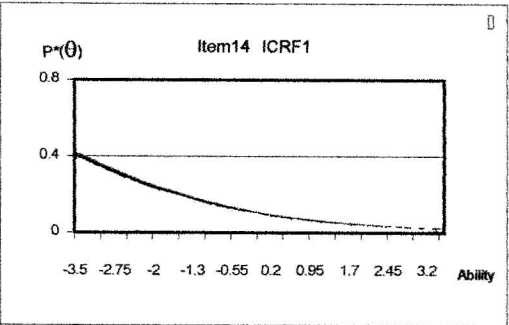
ลำดับชั้นที่ 5



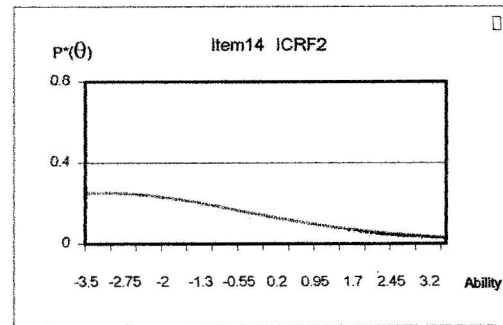
ลำดับชั้นที่ 6



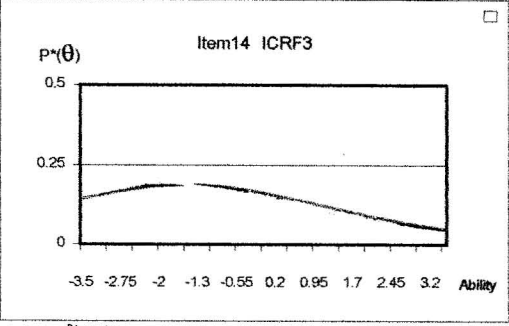
ข้อ 14 ลำดับชั้นที่ 1



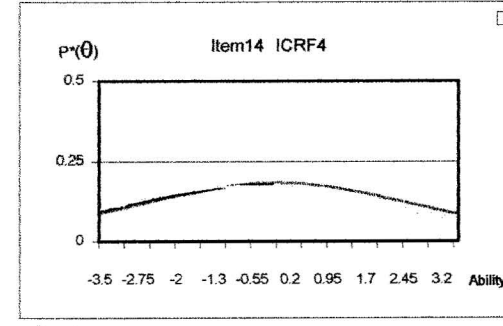
ลำดับชั้นที่ 2



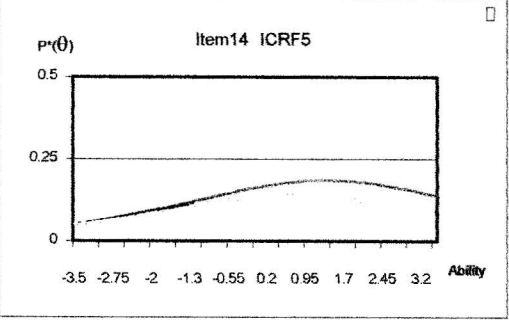
ลำดับชั้นที่ 3



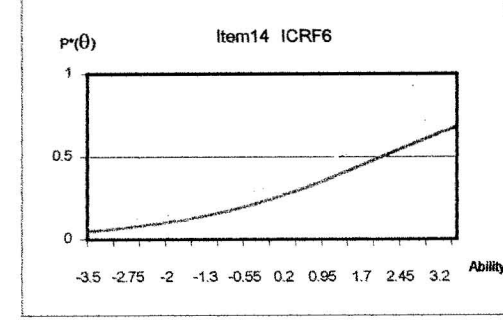
ลำดับชั้นที่ 4



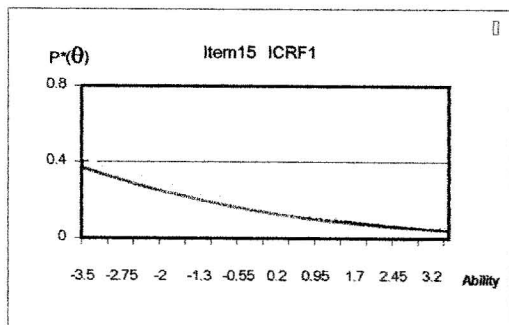
ลำดับชั้นที่ 5



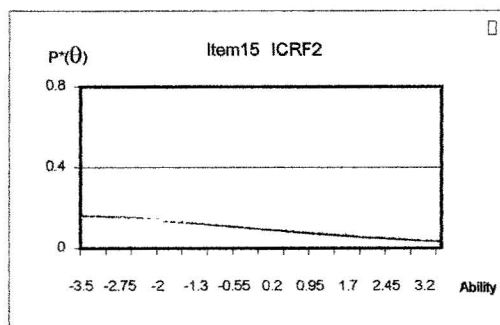
ลำดับชั้นที่ 6



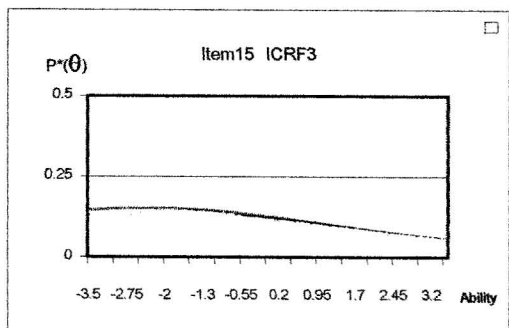
ข้อ 15 ลำดับขั้นที่ 1



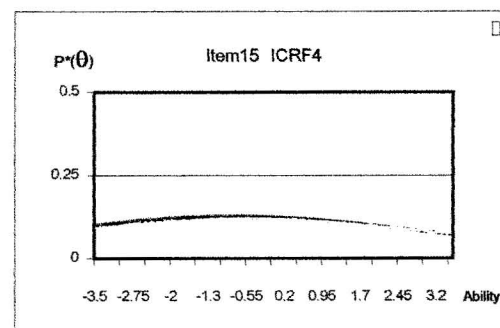
ลำดับขั้นที่ 2



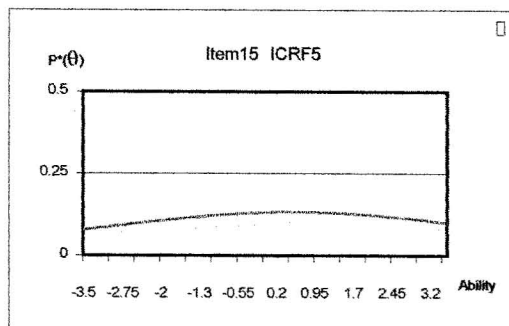
ลำดับขั้นที่ 3



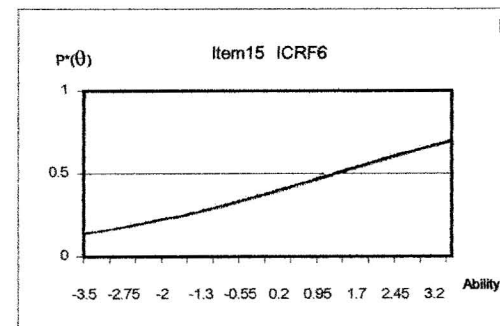
ลำดับขั้นที่ 4



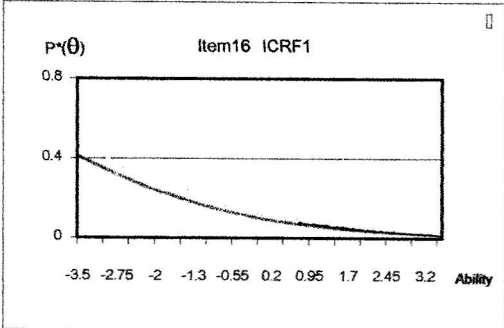
ลำดับขั้นที่ 5



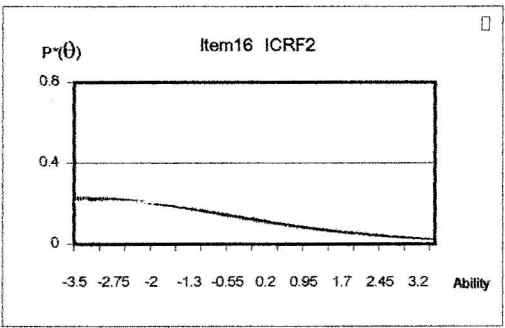
ลำดับขั้นที่ 6



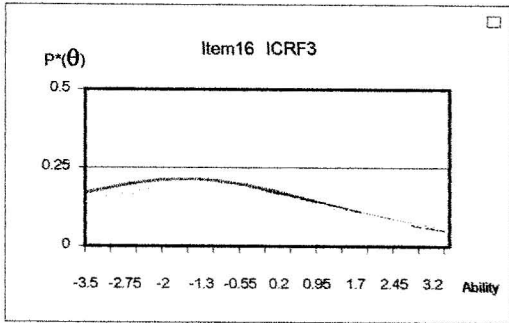
ข้อ 16 ลำดับชั้นที่ 1



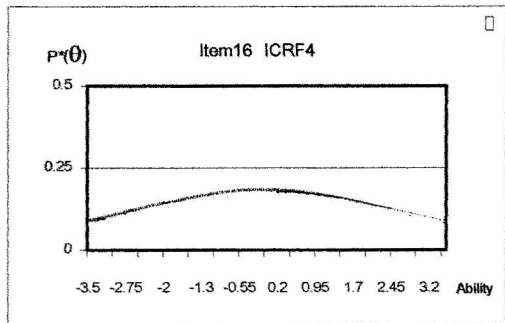
ลำดับชั้นที่ 2



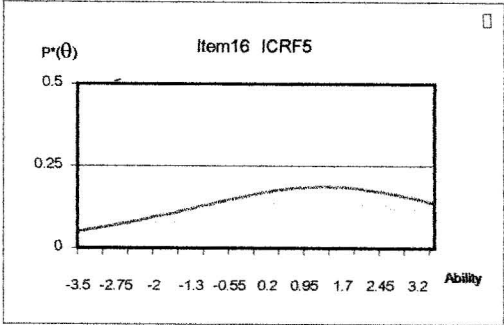
ลำดับชั้นที่ 3



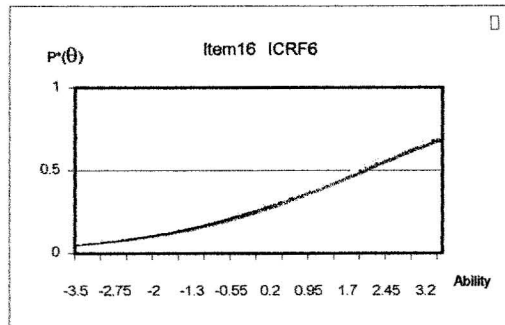
ลำดับชั้นที่ 4



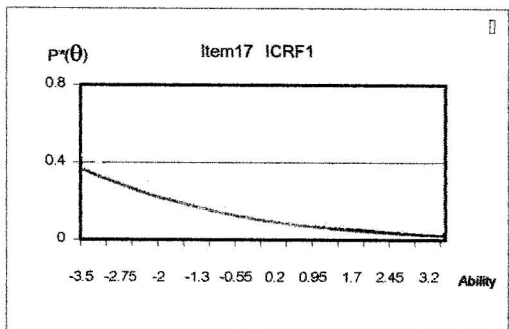
ลำดับชั้นที่ 5



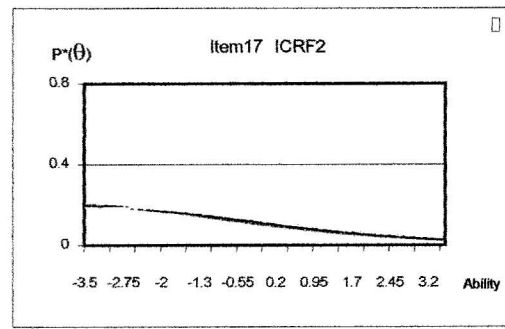
ลำดับชั้นที่ 6



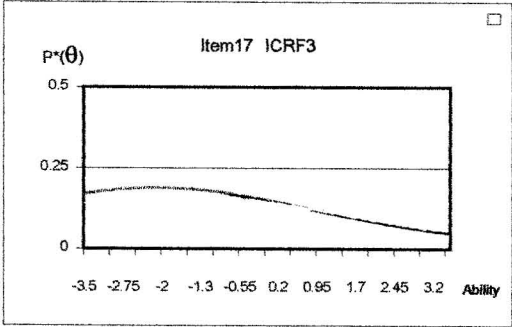
ข้อ 17 ลำดับชั้นที่ 1



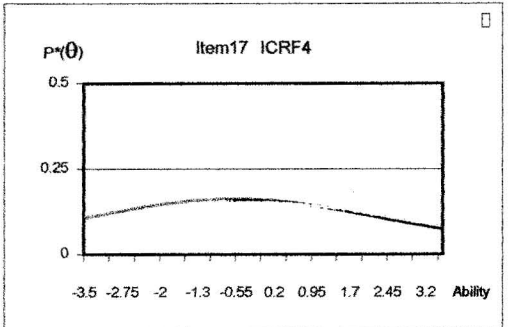
ลำดับชั้นที่ 2



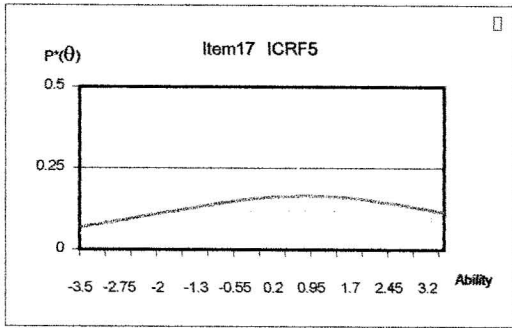
ลำดับชั้นที่ 3



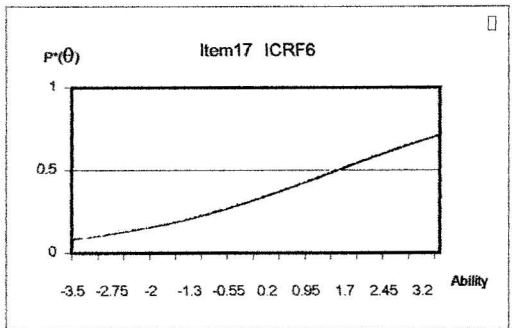
ลำดับชั้นที่ 4



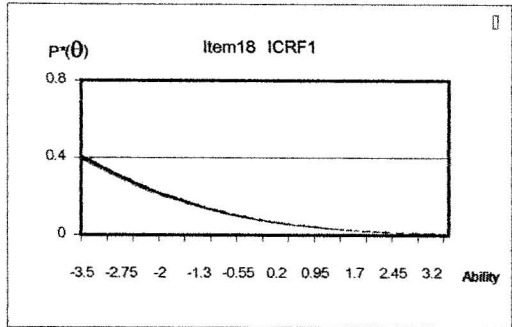
ลำดับชั้นที่ 5



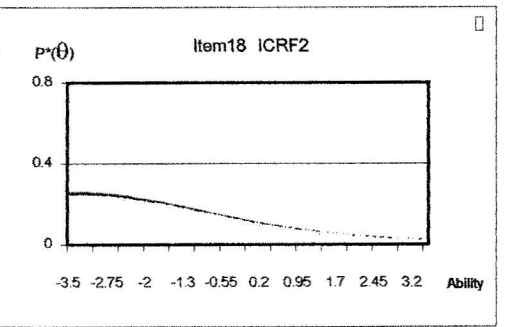
ลำดับชั้นที่ 6



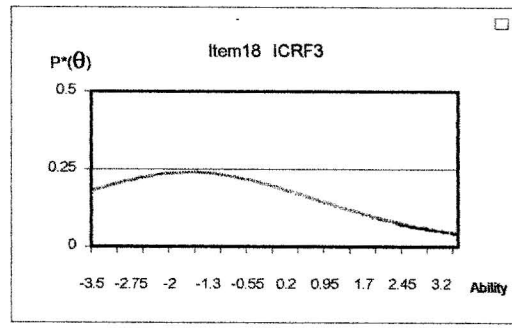
ข้อ 18 ลำดับชั้นที่ 1



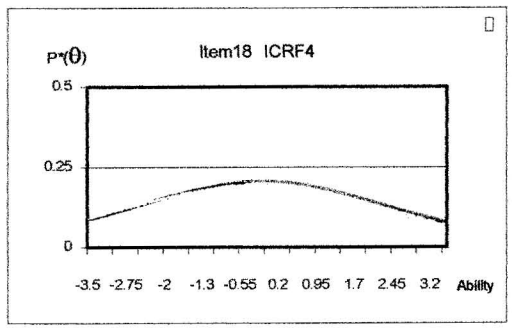
ลำดับชั้นที่ 2



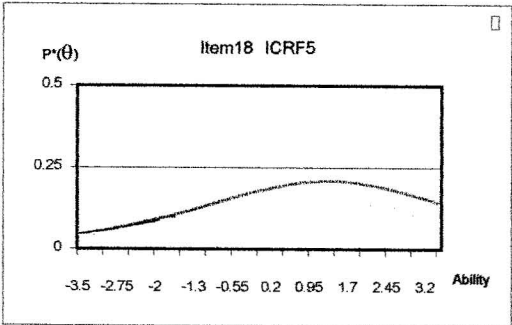
ลำดับชั้นที่ 3



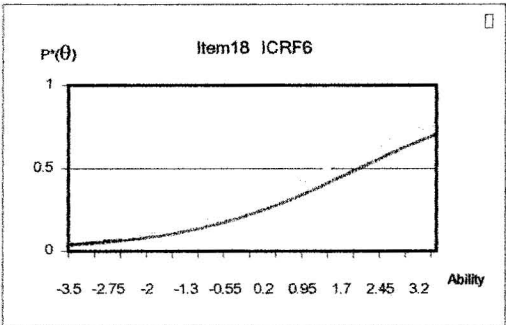
ลำดับชั้นที่ 4



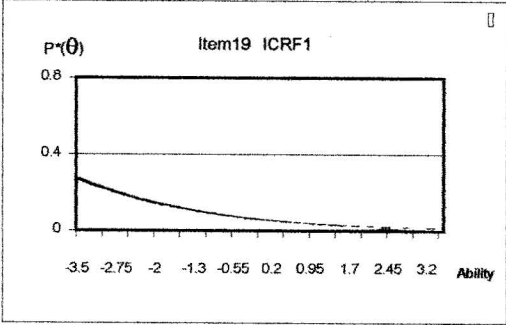
ลำดับชั้นที่ 5



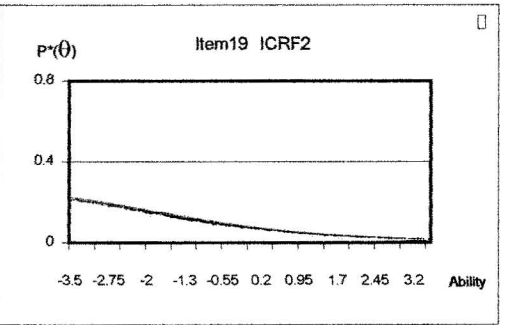
ลำดับชั้นที่ 6



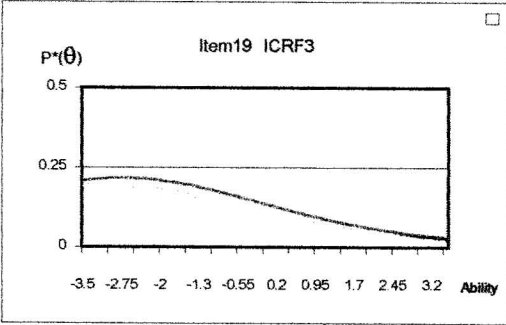
ข้อ 19 ลำดับชั้นที่ 1



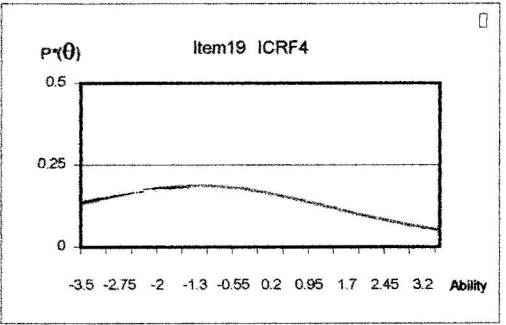
ลำดับชั้นที่ 2



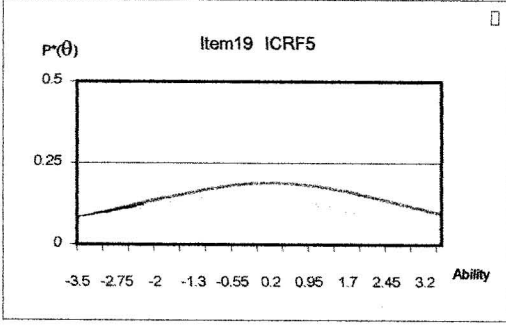
ลำดับชั้นที่ 3



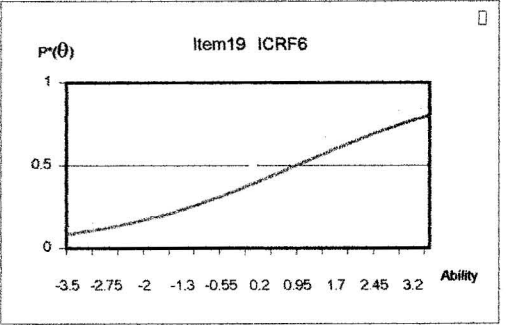
ลำดับชั้นที่ 4



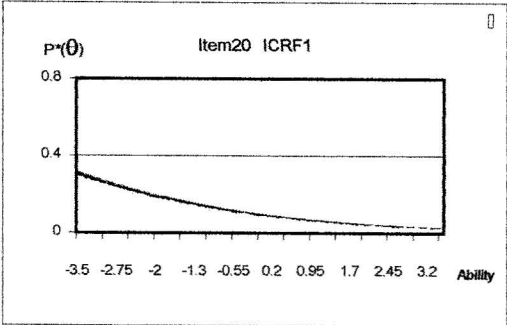
ลำดับชั้นที่ 5



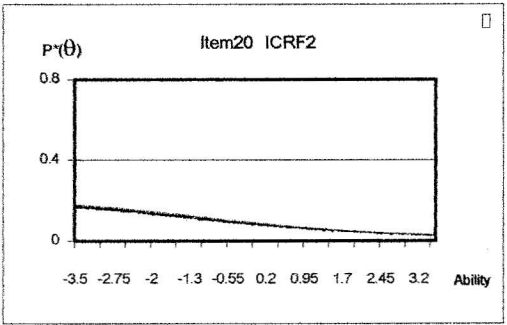
ลำดับชั้นที่ 6



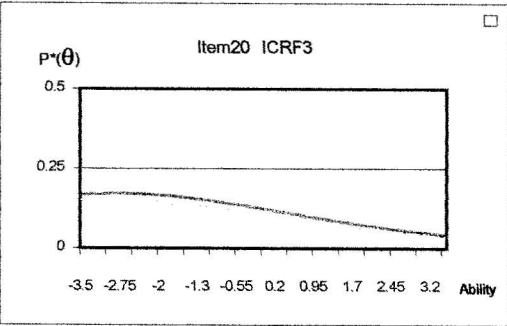
ข้อ 20 ลำดับชั้นที่ 1



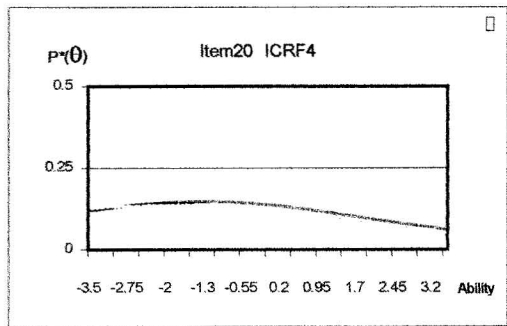
ลำดับชั้นที่ 2



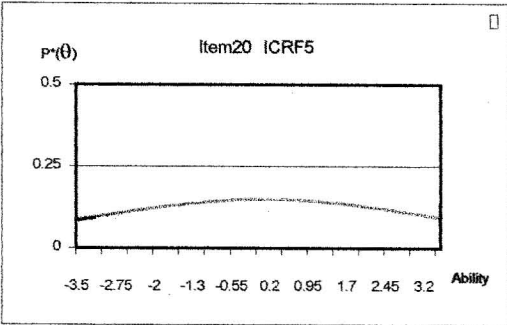
ลำดับชั้นที่ 3



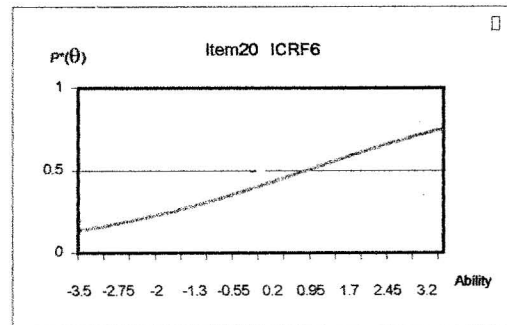
ลำดับชั้นที่ 4



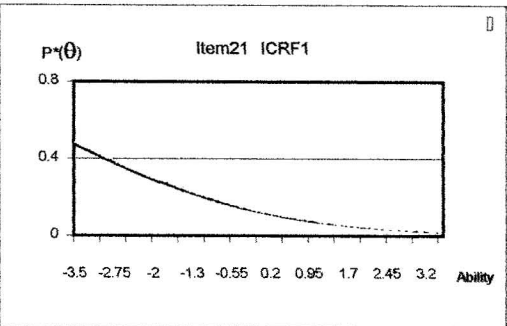
ลำดับชั้นที่ 5



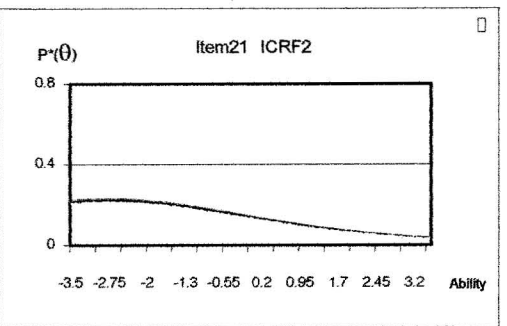
ลำดับชั้นที่ 6



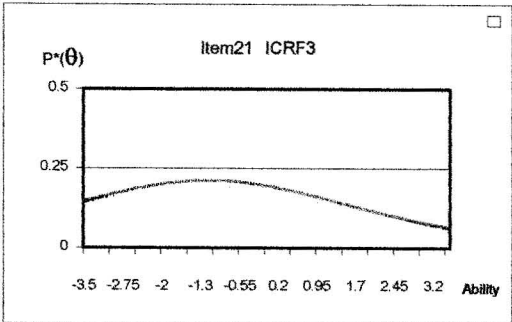
ข้อ 21 ลำดับชั้นที่ 1



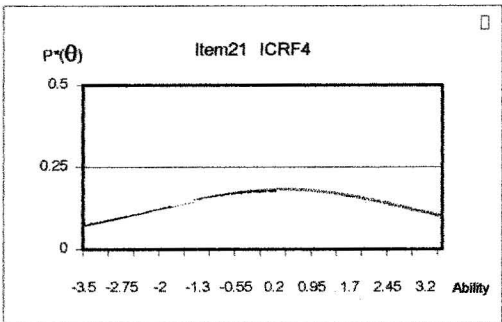
ลำดับชั้นที่ 2



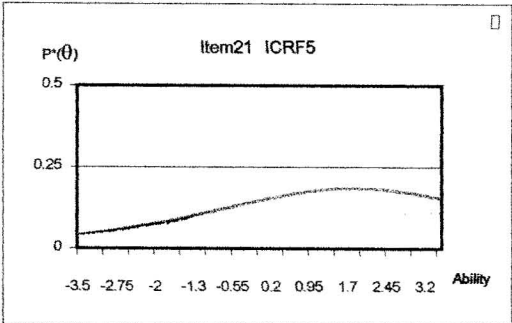
ลำดับชั้นที่ 3



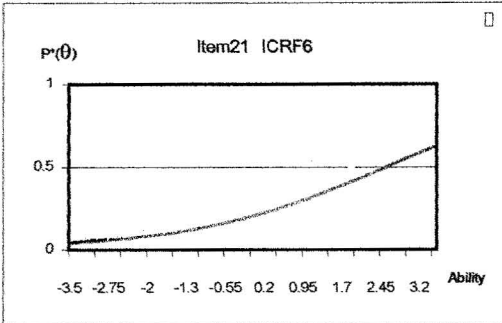
ลำดับชั้นที่ 4



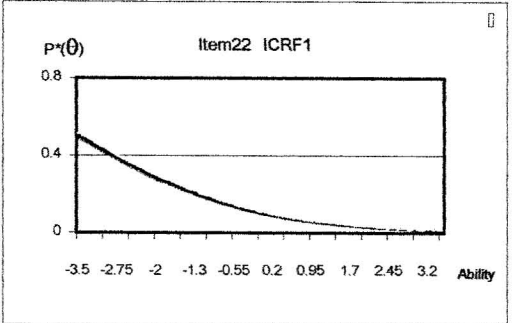
ลำดับชั้นที่ 5



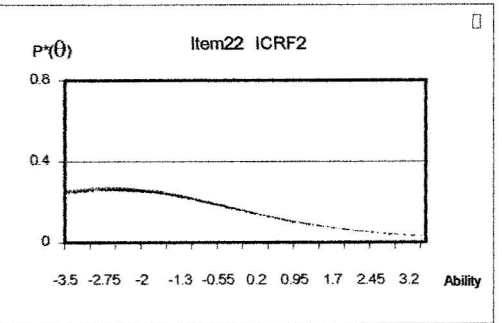
ลำดับชั้นที่ 6



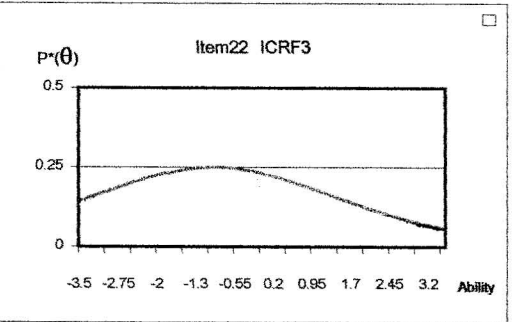
ข้อ 22 ลำดับชั้นที่ 1



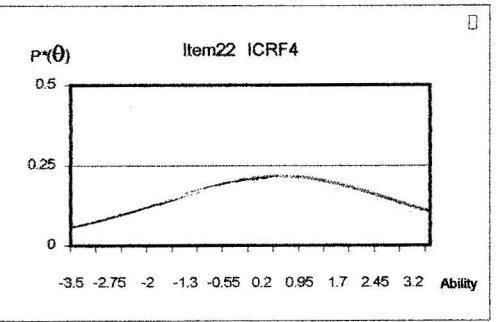
ลำดับชั้นที่ 2



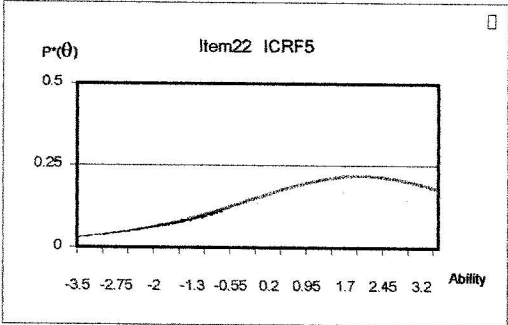
ลำดับชั้นที่ 3



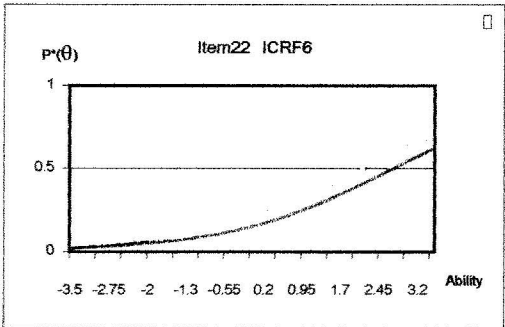
ลำดับชั้นที่ 4



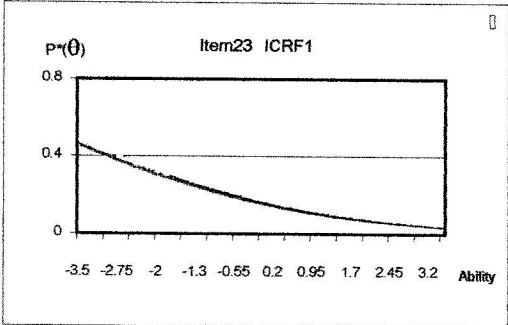
ลำดับชั้นที่ 5



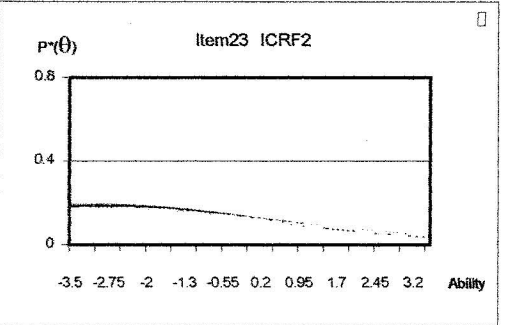
ลำดับชั้นที่ 6



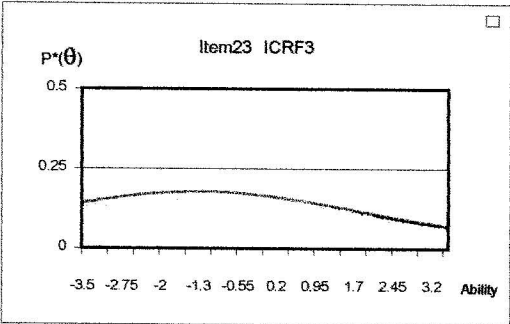
ข้อ 23 ลำดับชั้นที่ 1



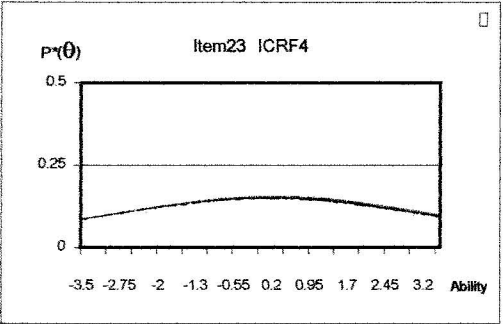
ลำดับชั้นที่ 2



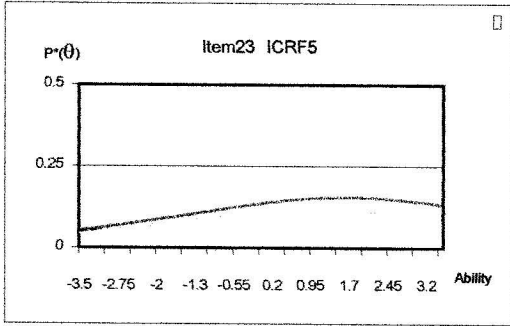
ลำดับชั้นที่ 3



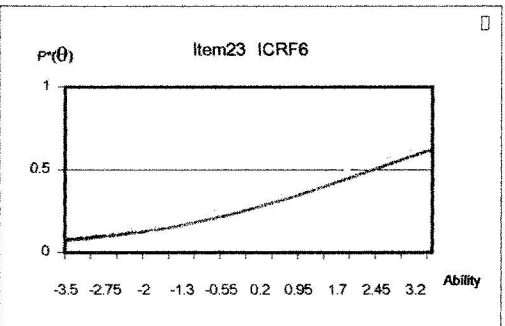
ลำดับชั้นที่ 4



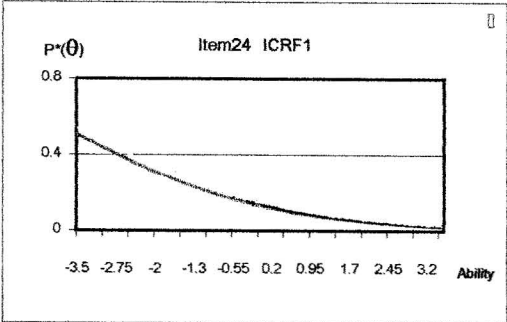
ลำดับชั้นที่ 5



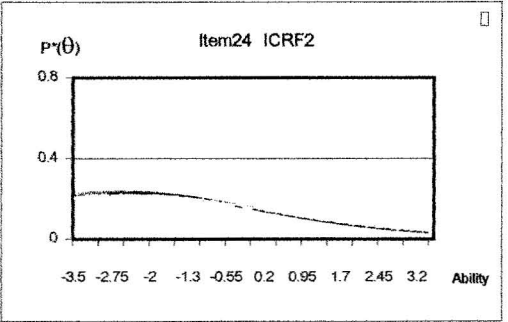
ลำดับชั้นที่ 6



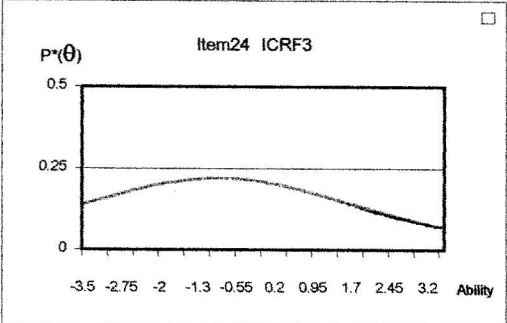
ข้อ 24 ลำดับขั้นที่ 1



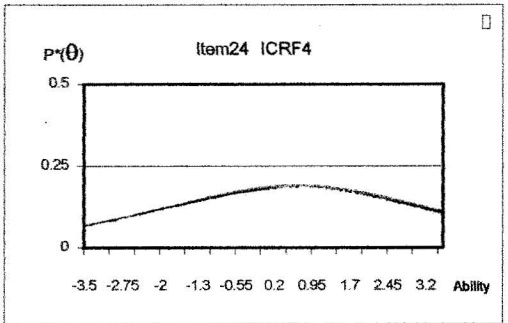
ลำดับขั้นที่ 2



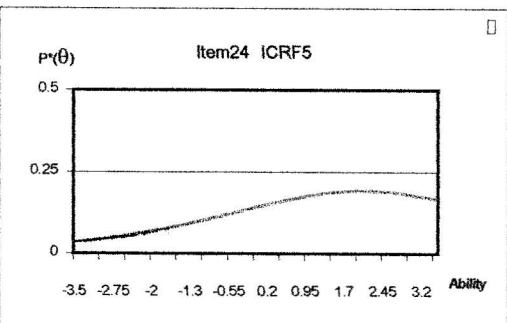
ลำดับขั้นที่ 3



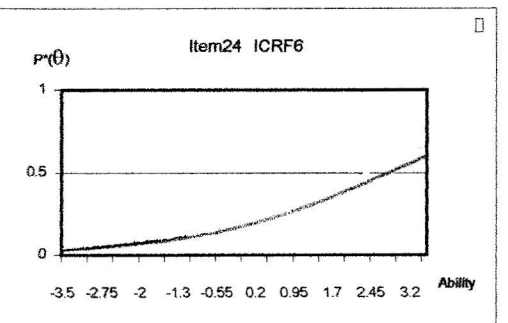
ลำดับขั้นที่ 4



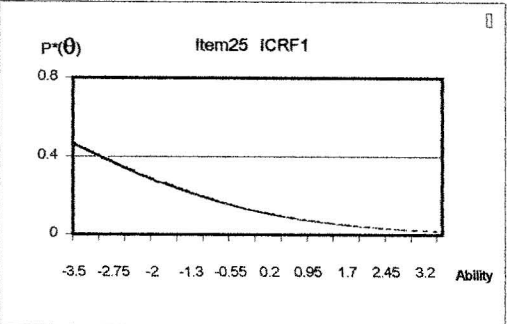
ลำดับขั้นที่ 5



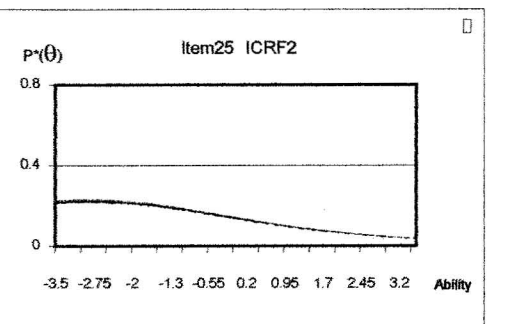
ลำดับขั้นที่ 6



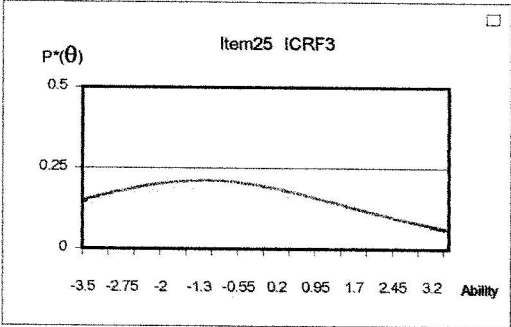
ข้อ 25 ลำดับขั้นที่ 1



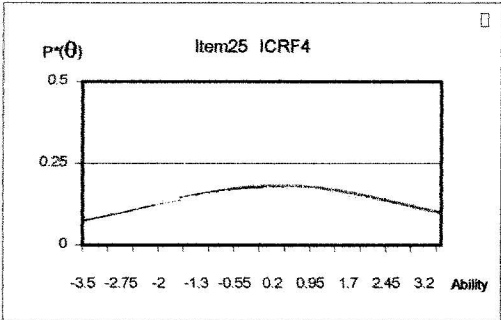
ลำดับขั้นที่ 2



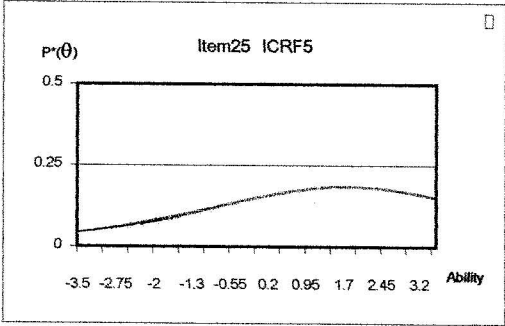
ลำดับชั้นที่ 3



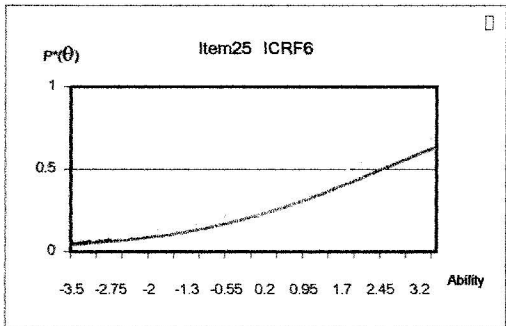
ลำดับชั้นที่ 4



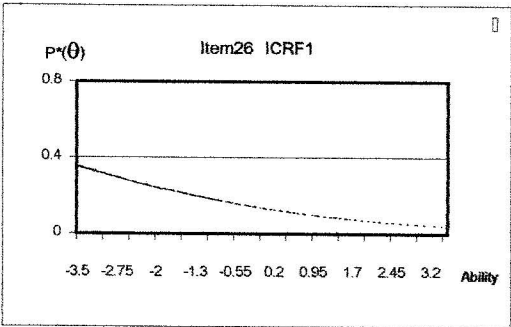
ลำดับชั้นที่ 5



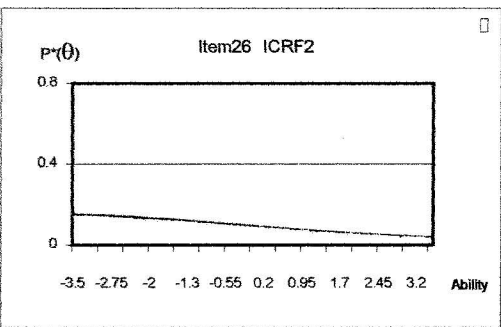
ลำดับชั้นที่ 6



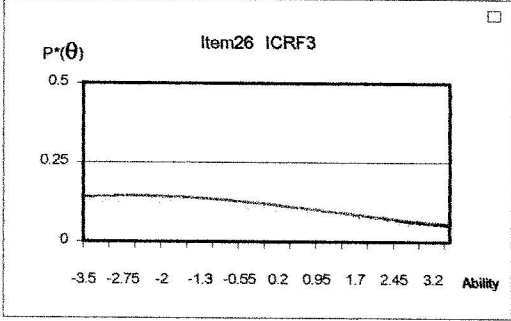
ข้อ 26 ลำดับชั้นที่ 1



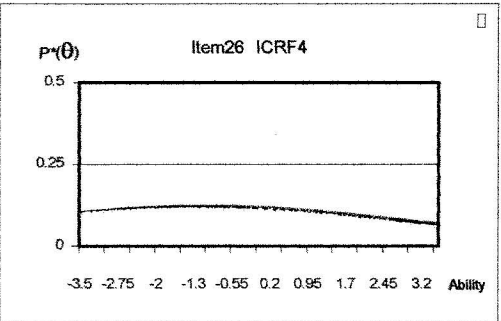
ลำดับชั้นที่ 2



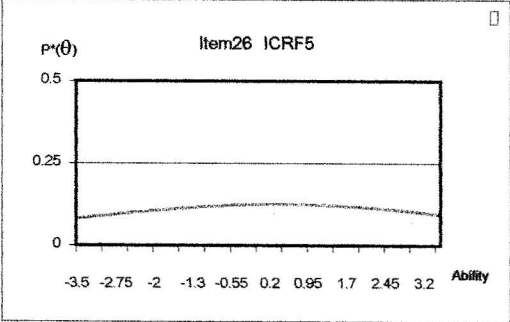
ลำดับชั้นที่ 3



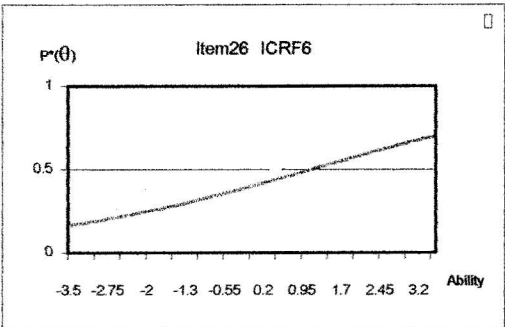
ลำดับชั้นที่ 4



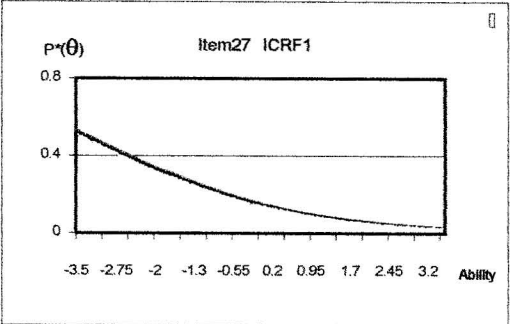
ลำดับชั้นที่ 5



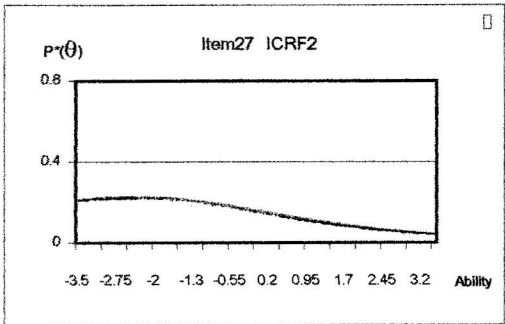
ลำดับชั้นที่ 6



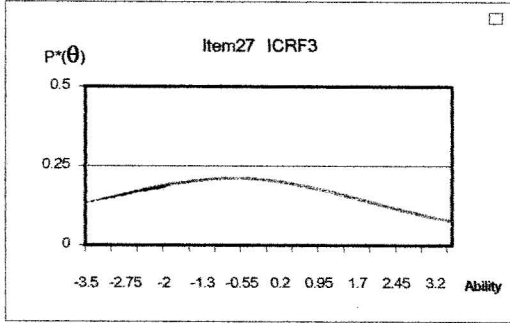
ข้อ 27 ลำดับชั้นที่ 1



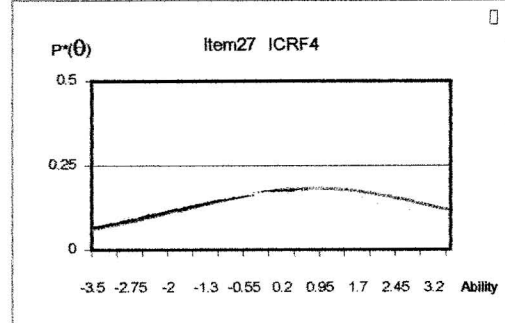
ลำดับชั้นที่ 2



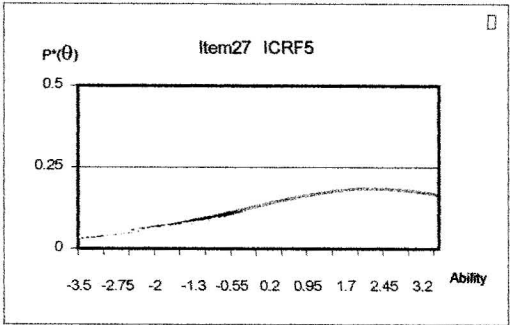
ลำดับชั้นที่ 3



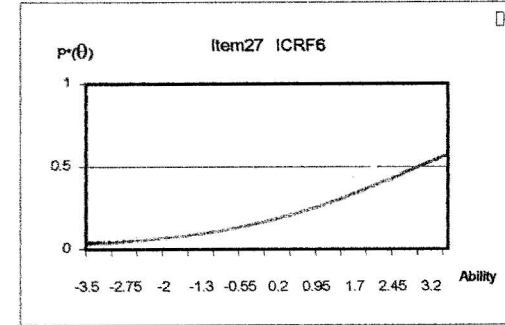
ลำดับชั้นที่ 4



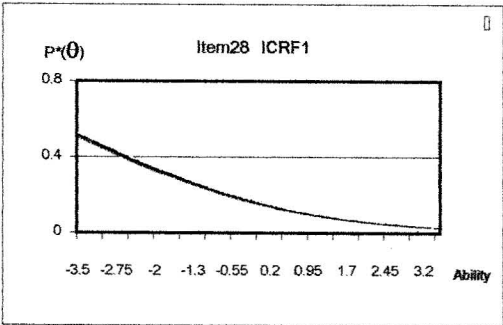
ลำดับชั้นที่ 5



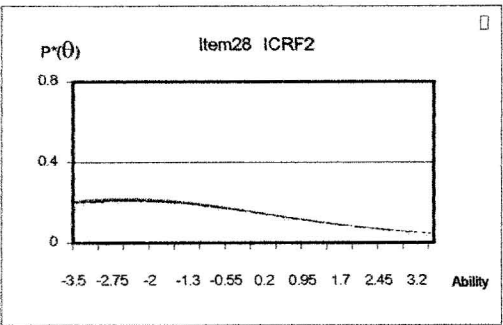
ลำดับชั้นที่ 6



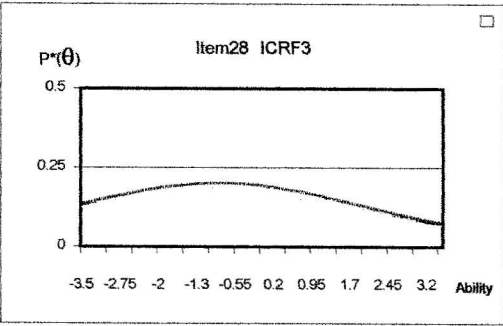
ข้อ 28 ลำดับขั้นที่ 1



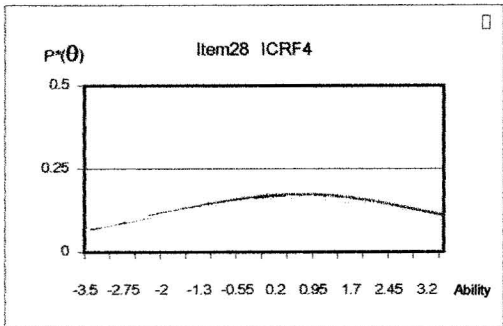
ลำดับขั้นที่ 2



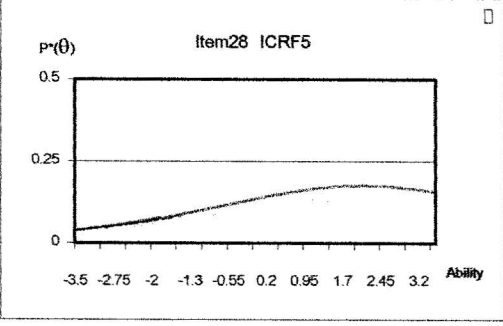
ลำดับขั้นที่ 3



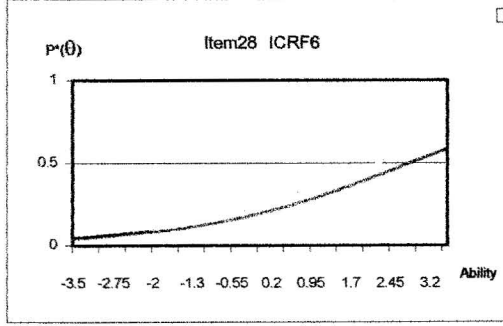
ลำดับขั้นที่ 4



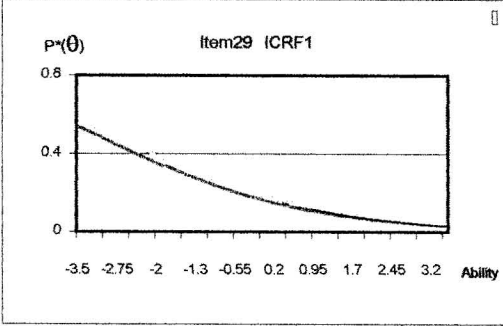
ลำดับขั้นที่ 5



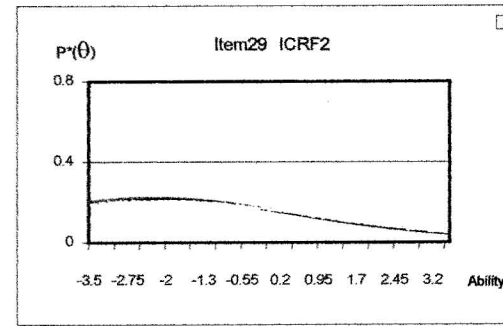
ลำดับขั้นที่ 6



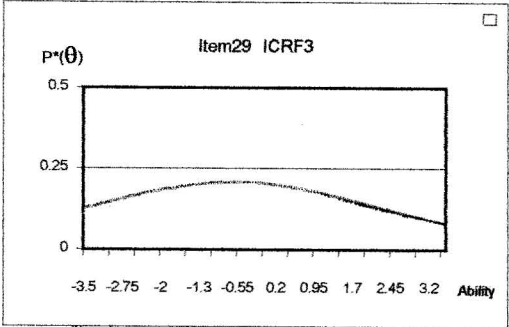
ข้อ 29 ลำดับขั้นที่ 1



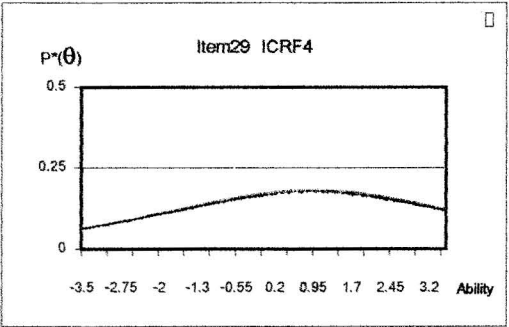
ลำดับขั้นที่ 2



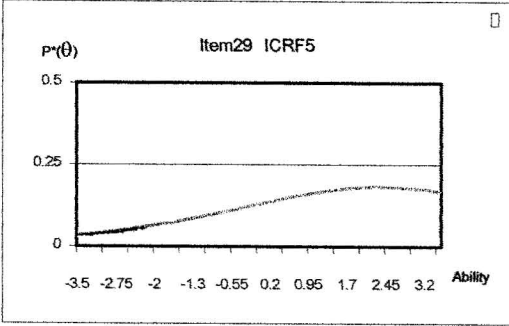
ลำดับชั้นที่ 3



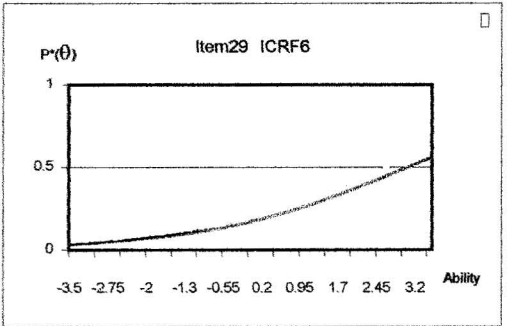
ลำดับชั้นที่ 4



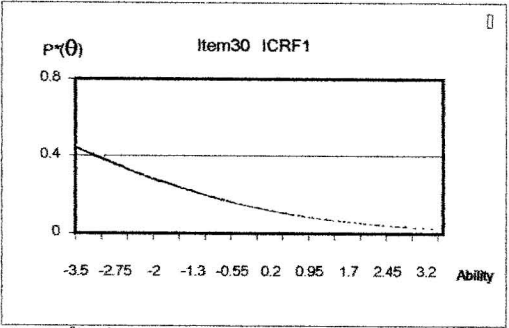
ลำดับชั้นที่ 5



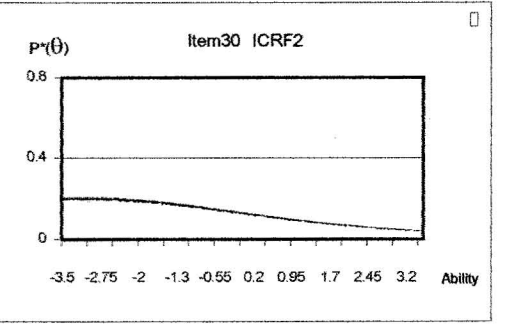
ลำดับชั้นที่ 6



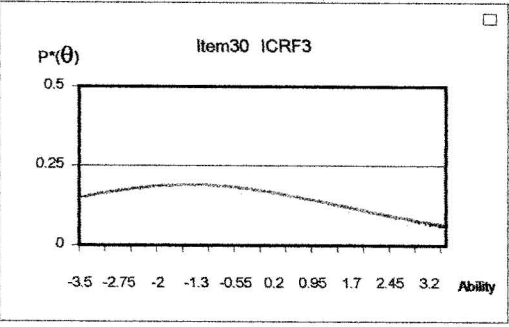
ข้อ 30 ลำดับชั้นที่ 1



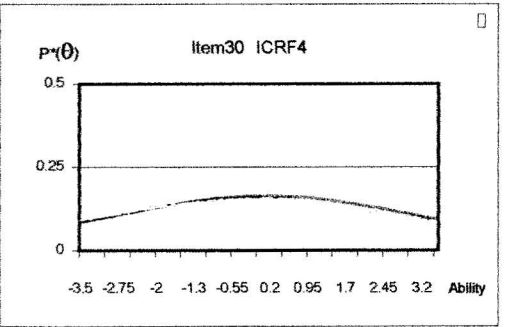
ลำดับชั้นที่ 2



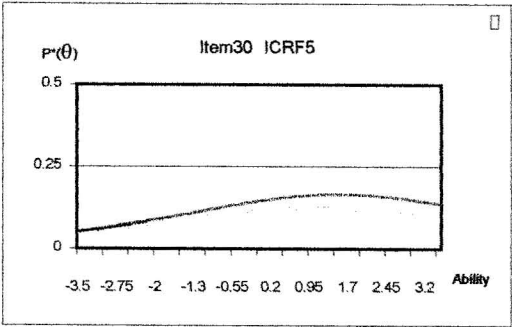
ลำดับชั้นที่ 3



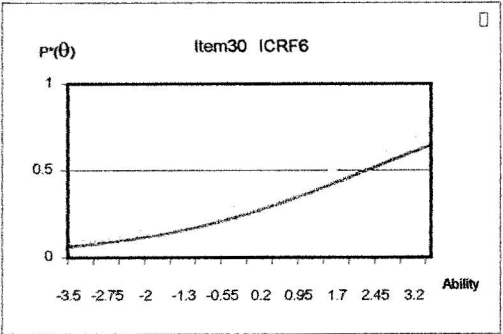
ลำดับชั้นที่ 4



ลำดับขั้นที่ 5



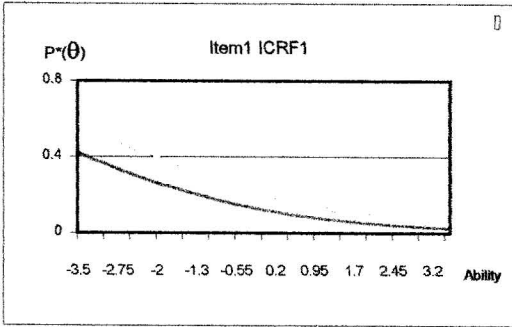
ลำดับขั้นที่ 6



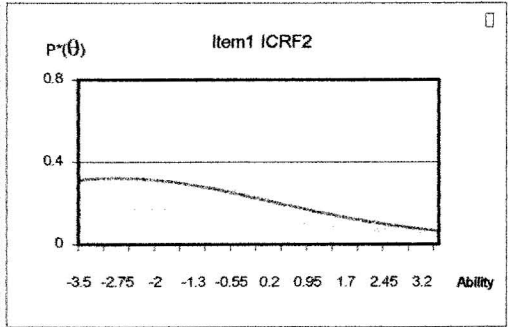
ฟังก์ชันผลการตอบลำดับชั้นข้อสอบ (ICRF) ที่วิเคราะห์ด้วยวิธี ICRF ในแบบทดสอบขนาด 20 ข้อ
เมื่อกลุ่มตัวอย่างเพศชายต่อเพศหญิง เท่ากับ 1,000 คน ต่อ 1,000 คน

ชาย หญิง

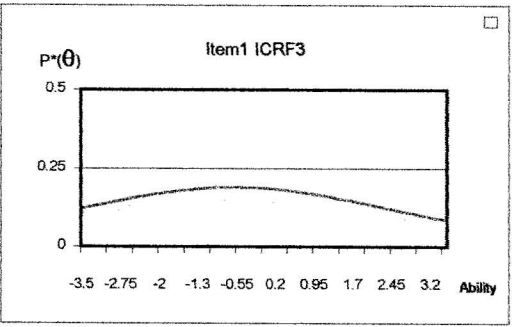
ข้อ 1 ลำดับชั้นที่ 1



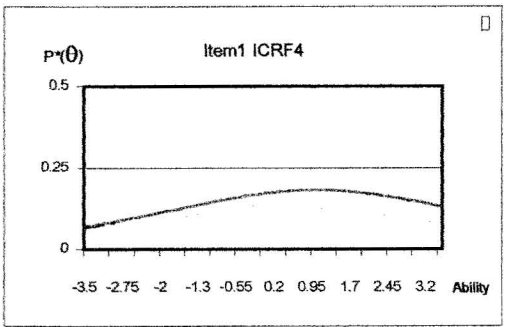
ลำดับชั้นที่ 2



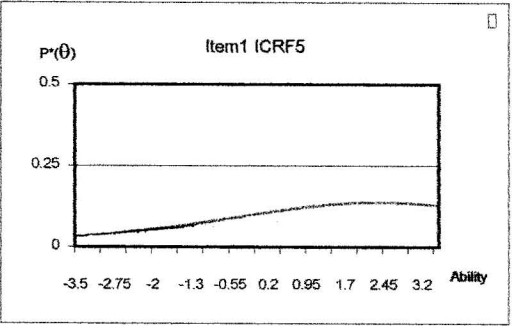
ลำดับชั้นที่ 3



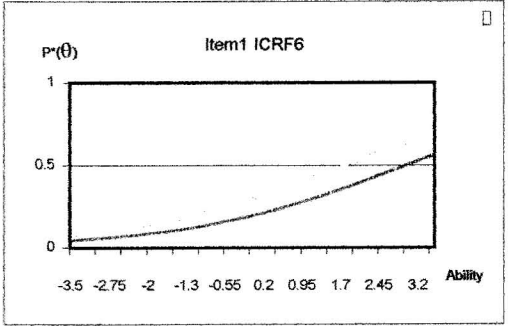
ลำดับชั้นที่ 4



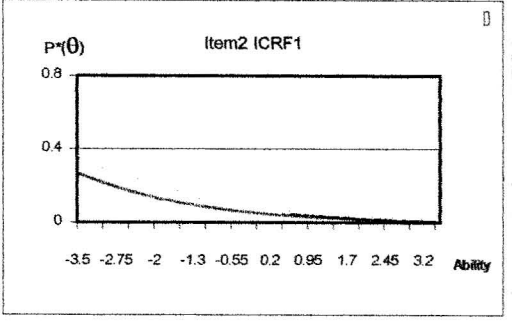
ลำดับชั้นที่ 5



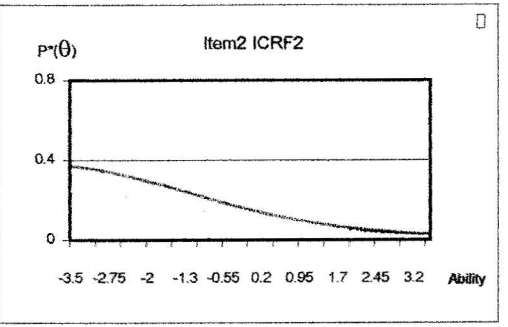
ลำดับชั้นที่ 6



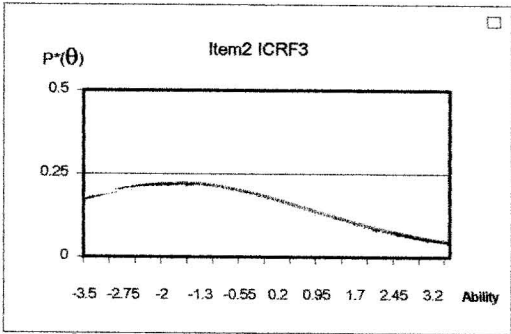
ข้อ 2 ลำดับชั้นที่ 1



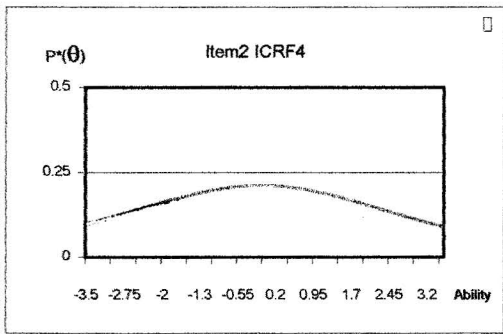
ลำดับชั้นที่ 2



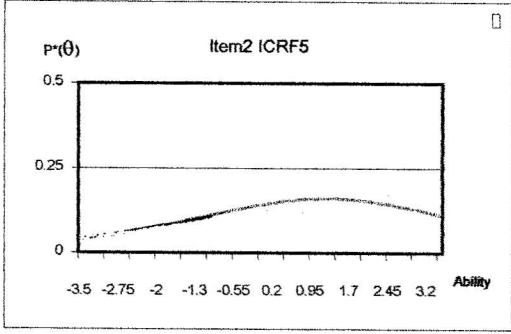
ลำดับขั้นที่ 3



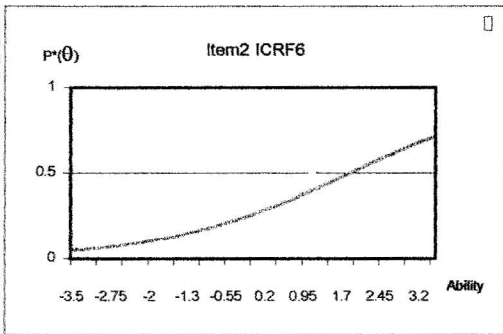
ลำดับขั้นที่ 4



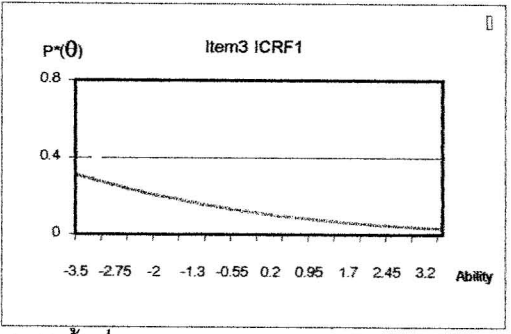
ลำดับขั้นที่ 5



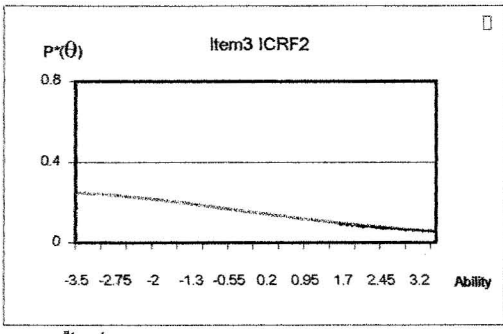
ลำดับขั้นที่ 6



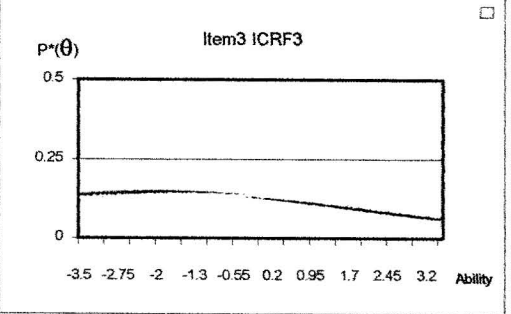
ข้อ 3 ลำดับขั้นที่ 1



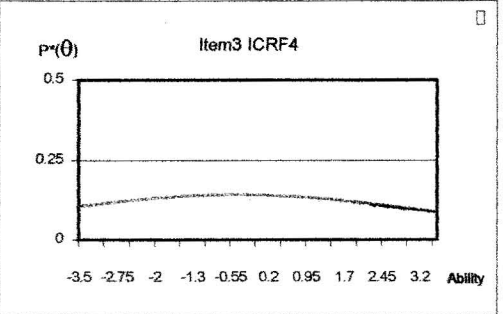
ลำดับขั้นที่ 2



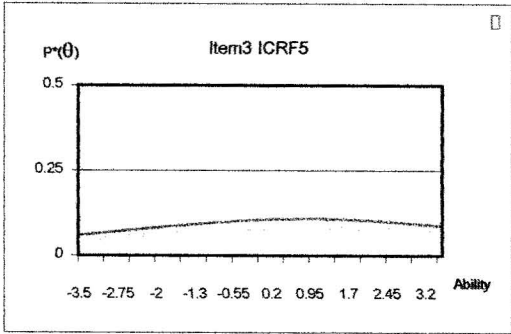
ลำดับขั้นที่ 3



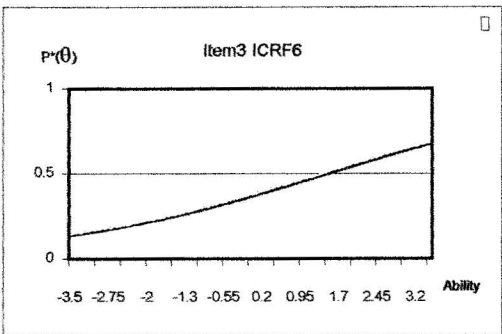
ลำดับขั้นที่ 4



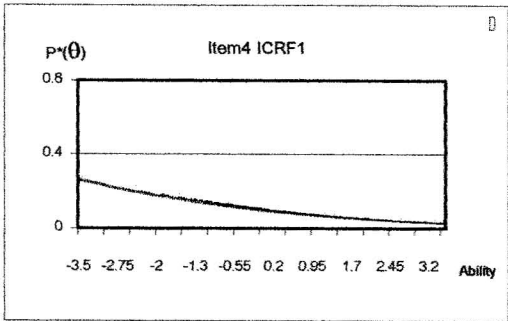
ลำดับขั้นที่ 5



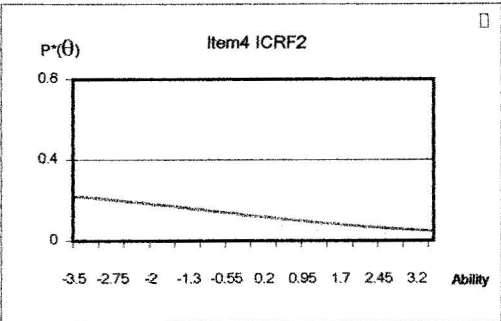
ลำดับขั้นที่ 6



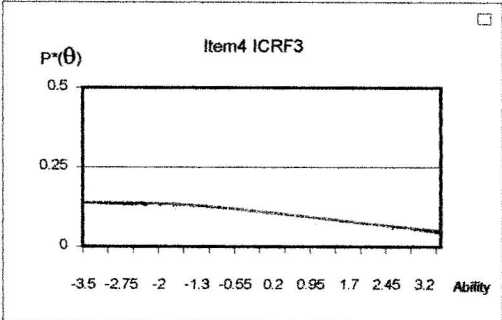
ข้อ 4 ลำดับขั้นที่ 1



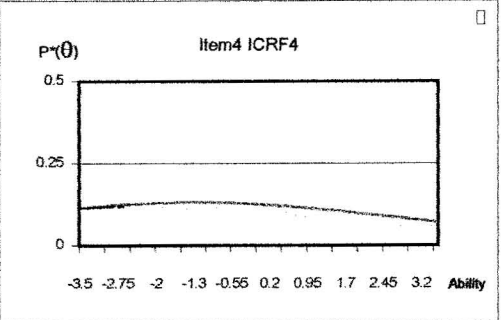
ลำดับขั้นที่ 2



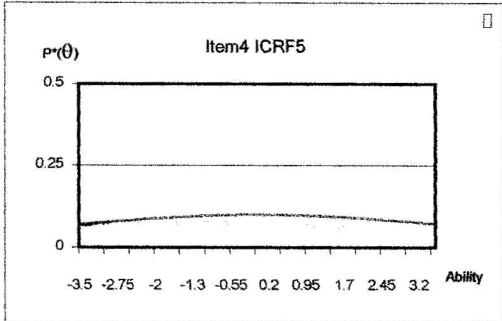
ลำดับขั้นที่ 3



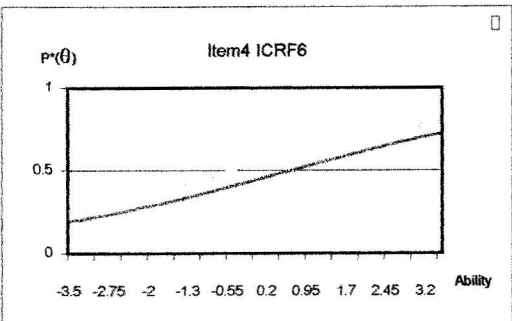
ลำดับขั้นที่ 4



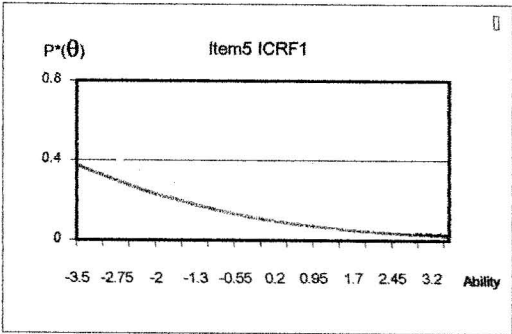
ลำดับขั้นที่ 5



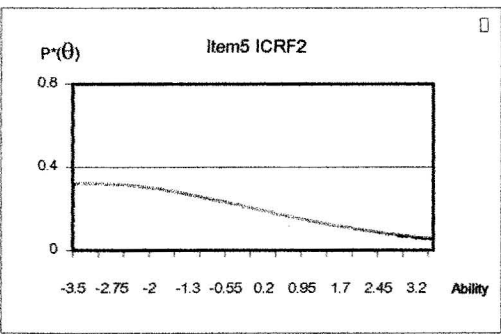
ลำดับขั้นที่ 6



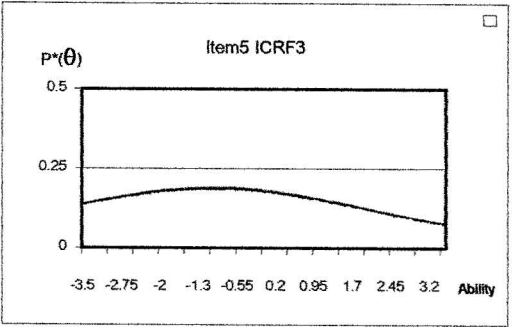
ข้อ 5 ลำดับขั้นที่ 1



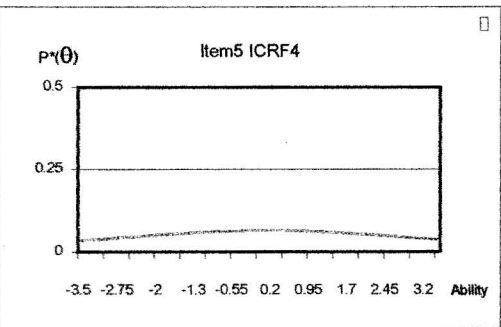
ลำดับขั้นที่ 2



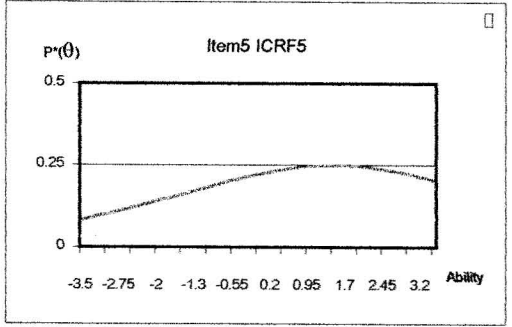
ลำดับขั้นที่ 3



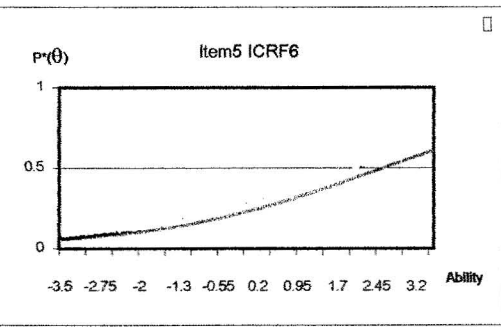
ลำดับขั้นที่ 4



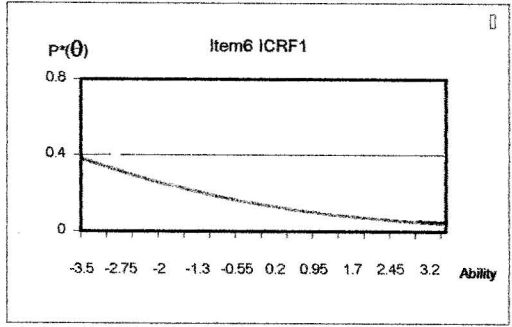
ลำดับขั้นที่ 5



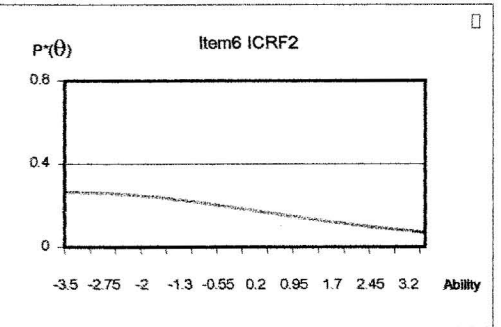
ลำดับขั้นที่ 6



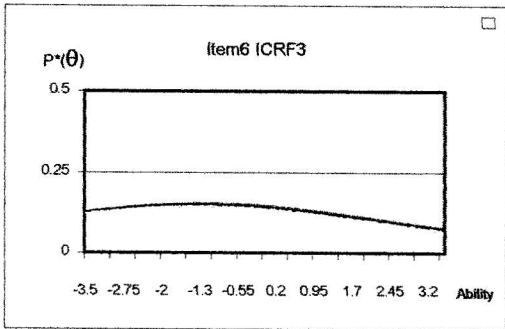
ข้อ 6 ลำดับขั้นที่ 1



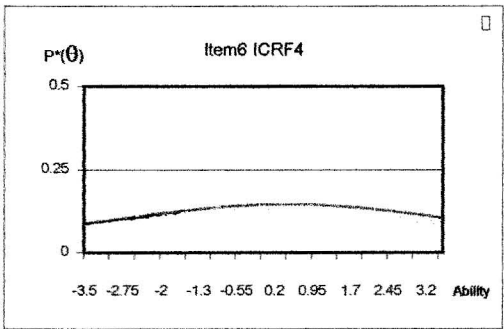
ลำดับขั้นที่ 2



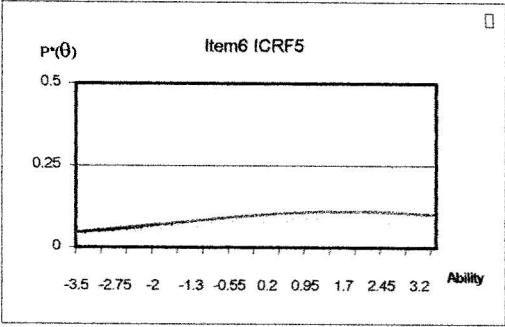
ลำดับชั้นที่ 3



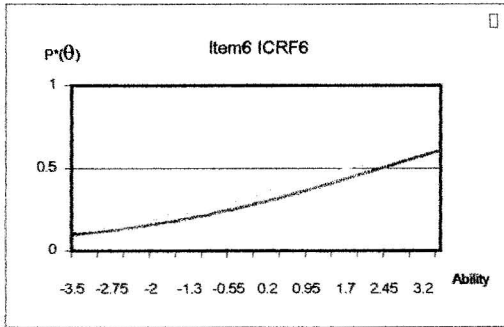
ลำดับชั้นที่ 4



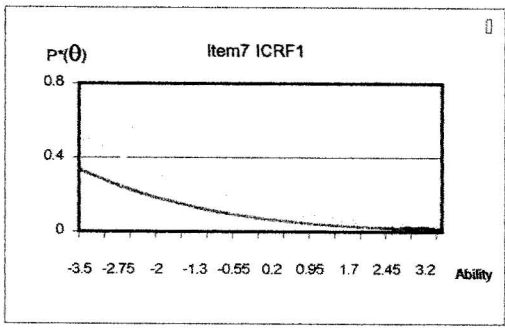
ลำดับชั้นที่ 5



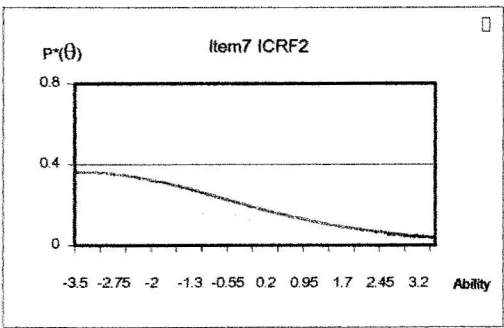
ลำดับชั้นที่ 6



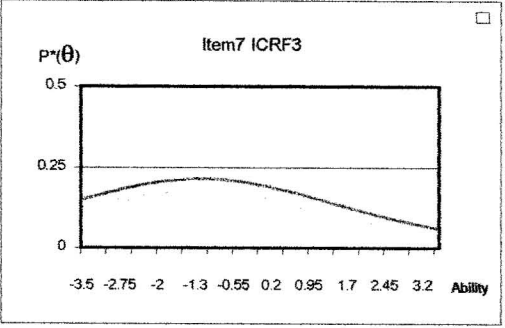
ข้อ 7 ลำดับชั้นที่ 1



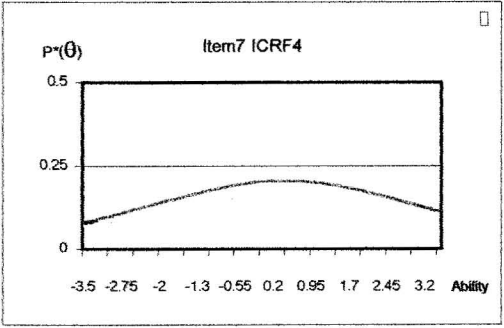
ลำดับชั้นที่ 2



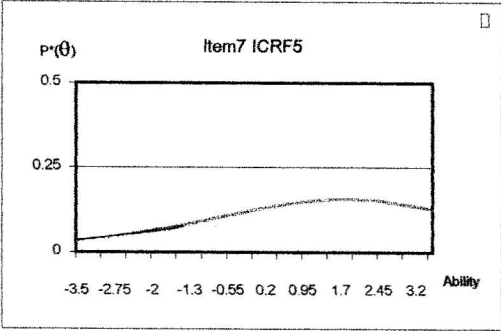
ลำดับชั้นที่ 3



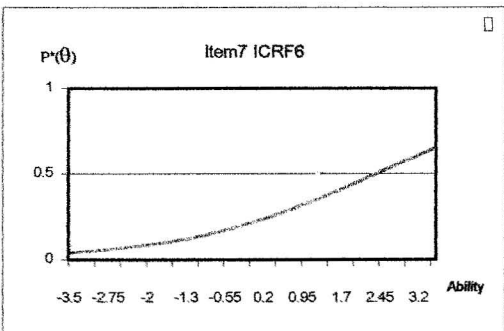
ลำดับชั้นที่ 4



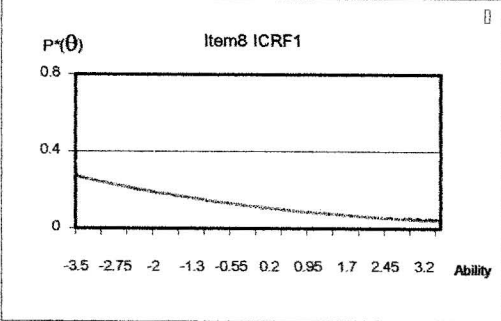
ลำดับชั้นที่ 5



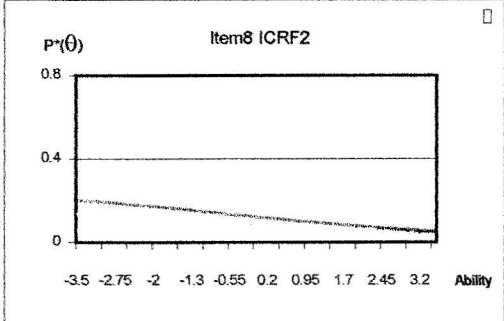
ลำดับชั้นที่ 6



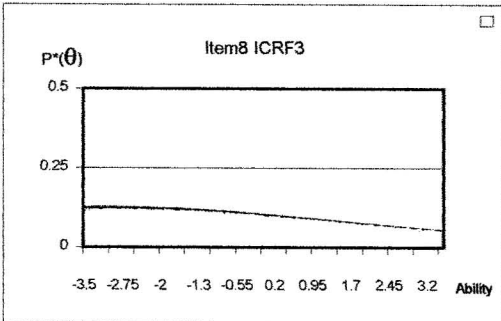
ข้อ 8 ลำดับชั้นที่ 1



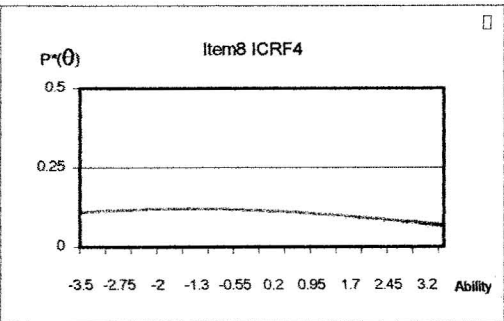
ลำดับชั้นที่ 2



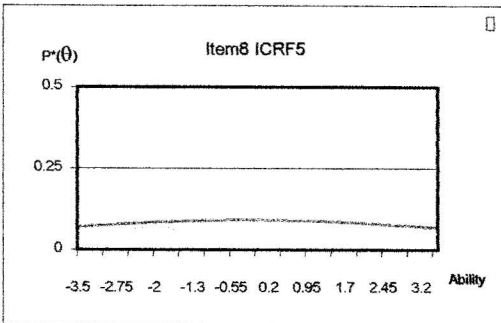
ลำดับชั้นที่ 3



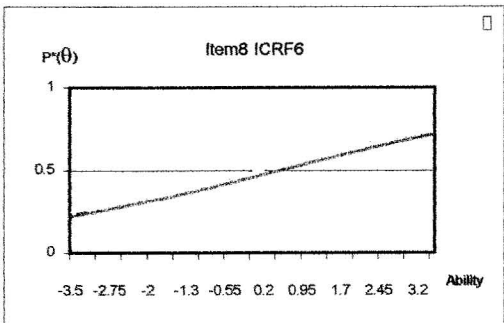
ลำดับชั้นที่ 4



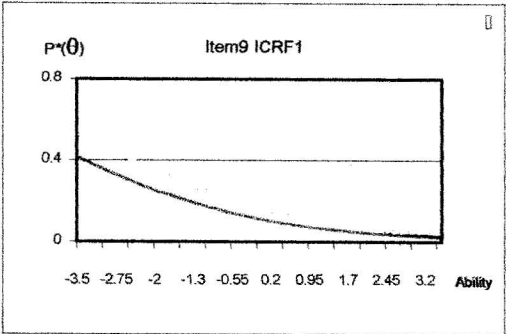
ลำดับชั้นที่ 5



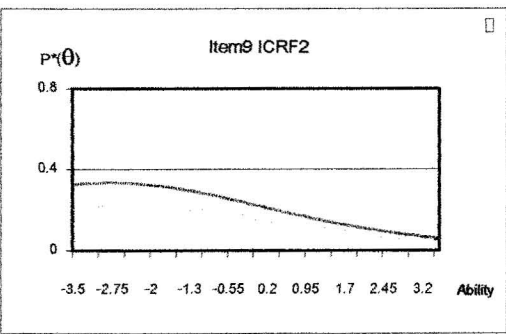
ลำดับชั้นที่ 6



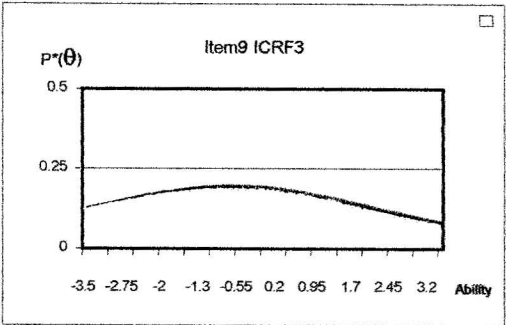
ข้อ 9 ลำดับขั้นที่ 1



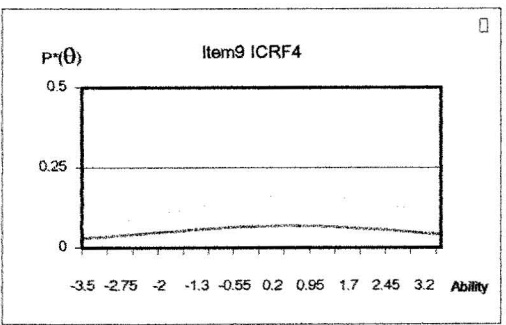
ลำดับขั้นที่ 2



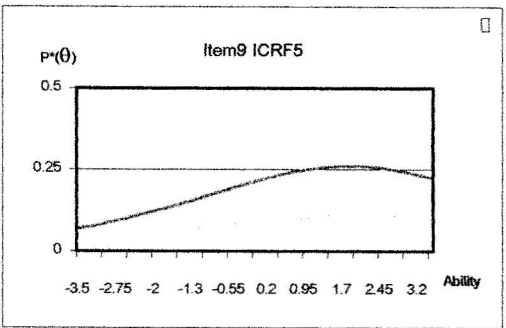
ลำดับขั้นที่ 3



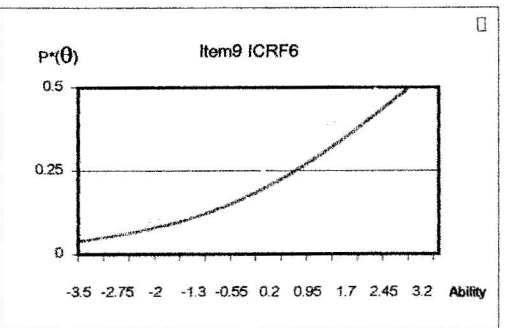
ลำดับขั้นที่ 4



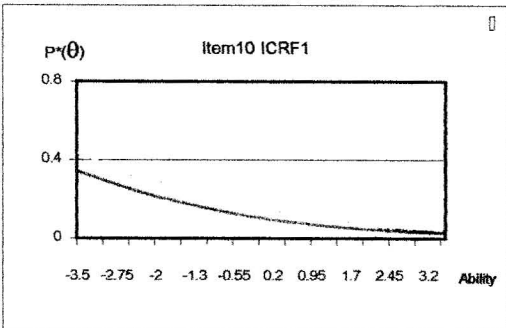
ลำดับขั้นที่ 5



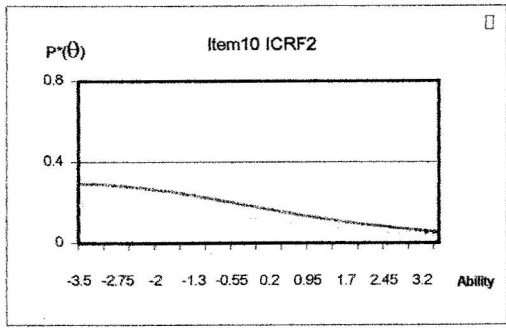
ลำดับขั้นที่ 6



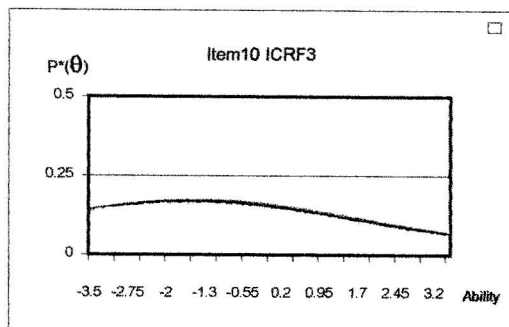
ข้อ 10 ลำดับขั้นที่ 1



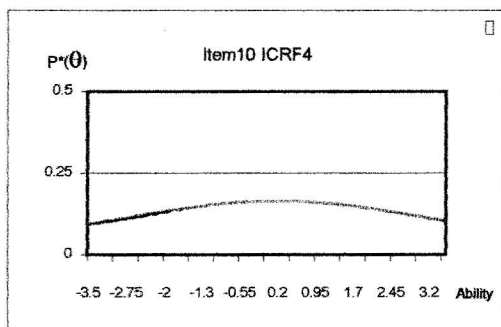
ลำดับขั้นที่ 2



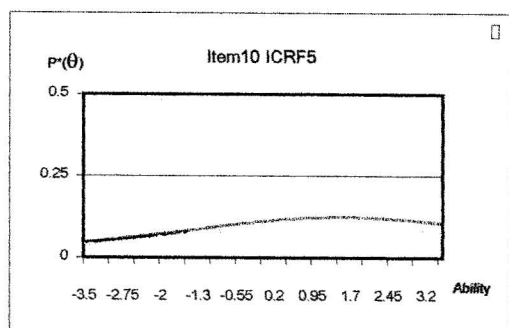
ลำดับชั้นที่ 3



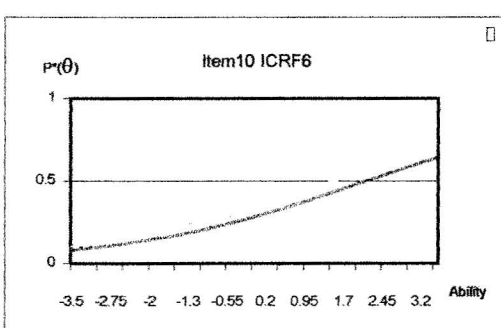
ลำดับชั้นที่ 4



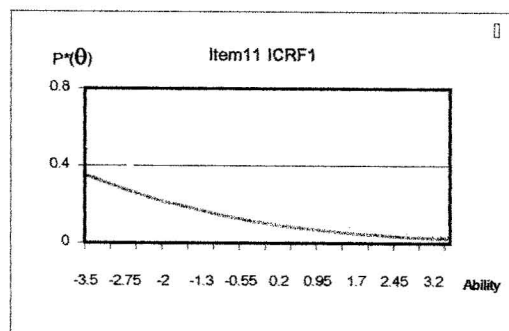
ลำดับชั้นที่ 5



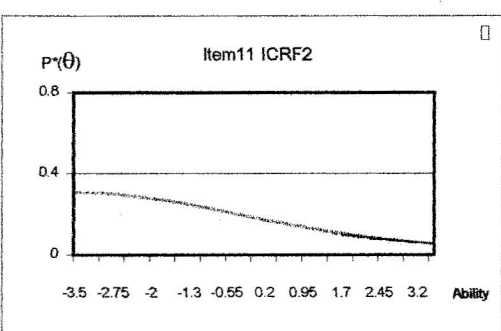
ลำดับชั้นที่ 6



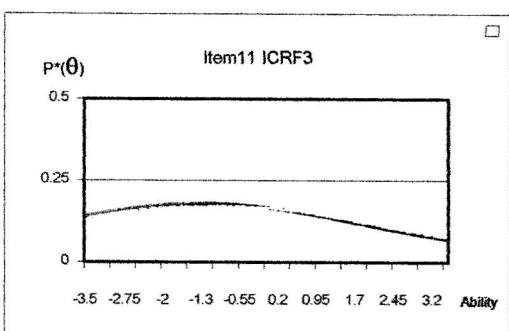
ข้อ 11 ลำดับชั้นที่ 1



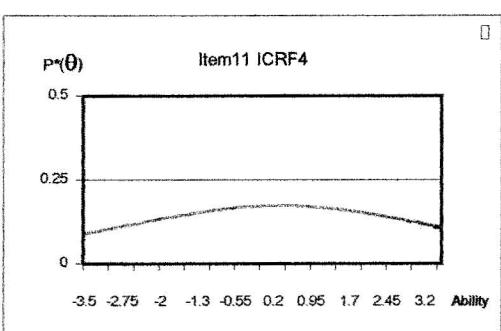
ลำดับชั้นที่ 2



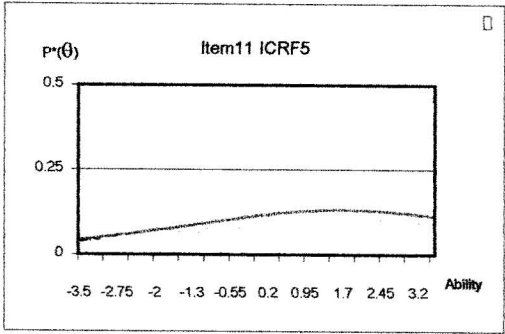
ลำดับชั้นที่ 3



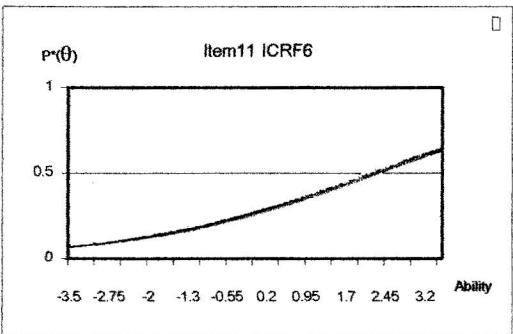
ลำดับชั้นที่ 4



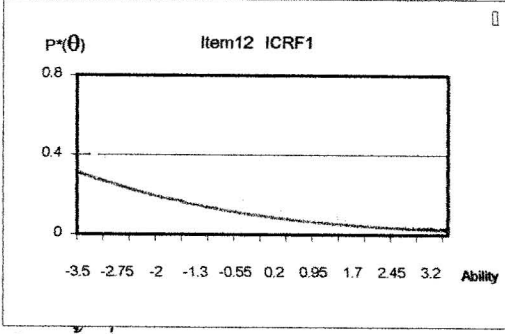
ลำดับชั้นที่ 5



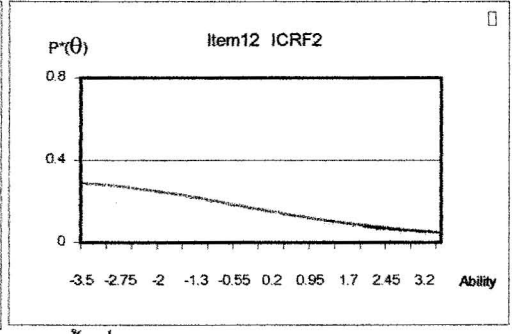
ลำดับชั้นที่ 6



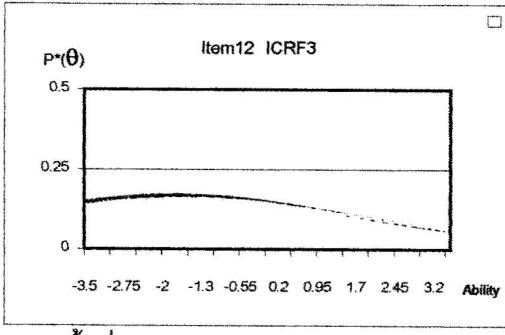
ข้อ 12 ลำดับชั้นที่ 1



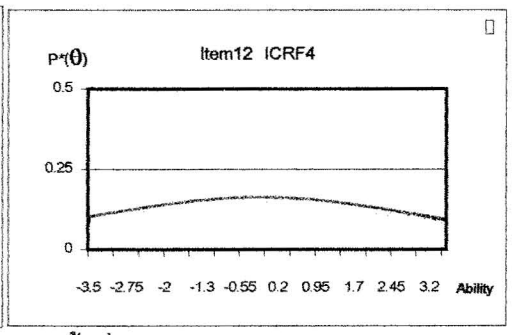
ลำดับชั้นที่ 2



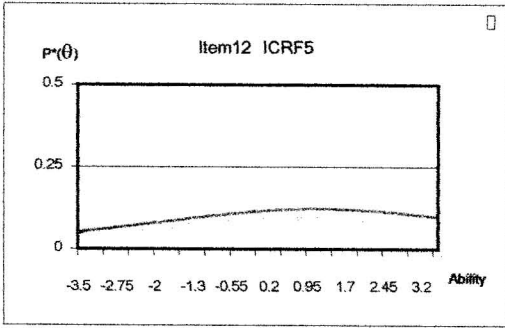
ลำดับชั้นที่ 3



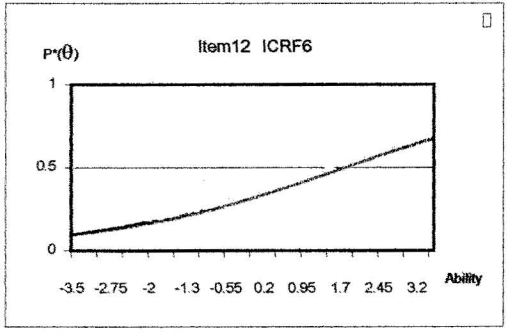
ลำดับชั้นที่ 4



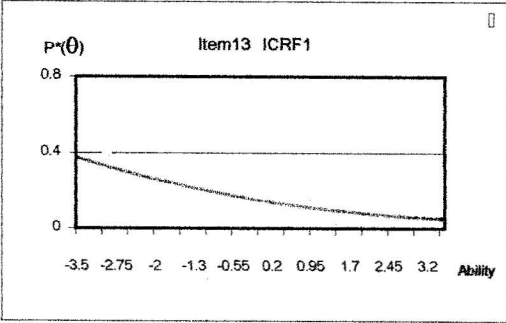
ลำดับชั้นที่ 5



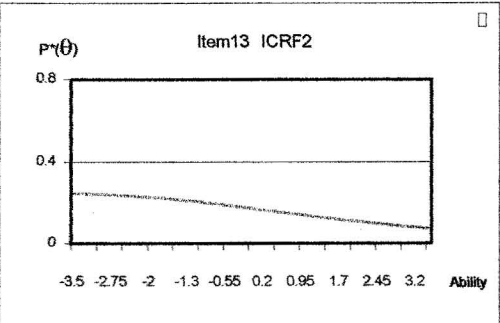
ลำดับชั้นที่ 6



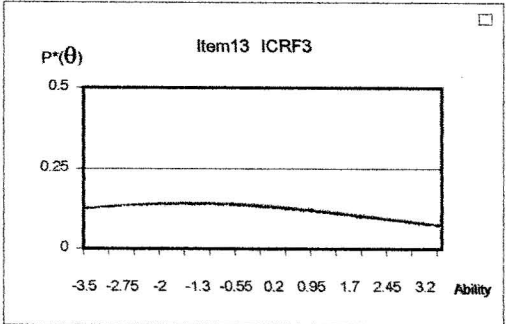
ข้อ 13 ลำดับขั้นที่ 1



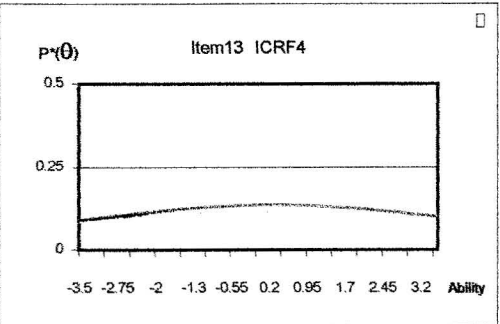
ลำดับขั้นที่ 2



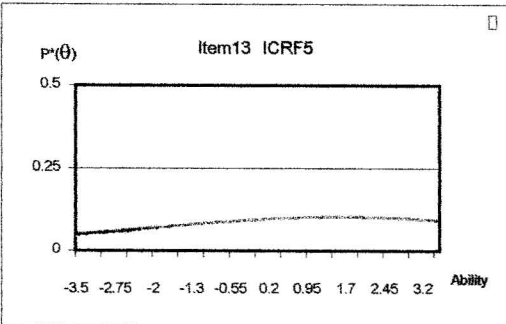
ลำดับขั้นที่ 3



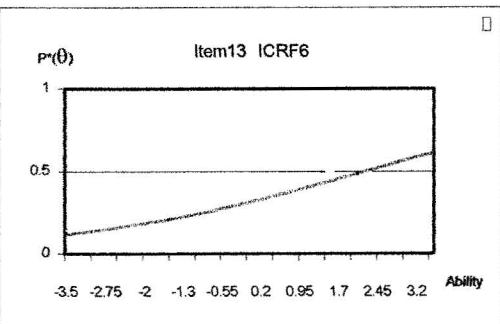
ลำดับขั้นที่ 4



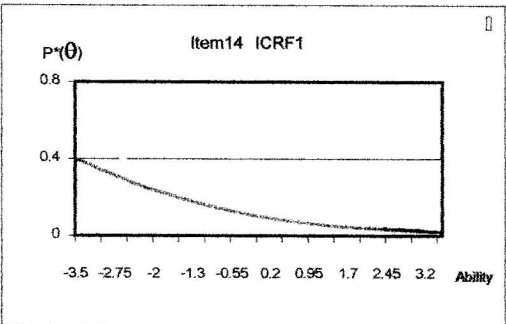
ลำดับขั้นที่ 5



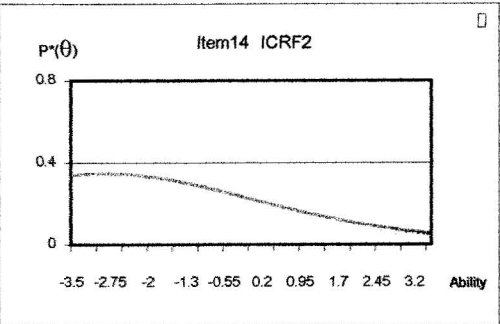
ลำดับขั้นที่ 6



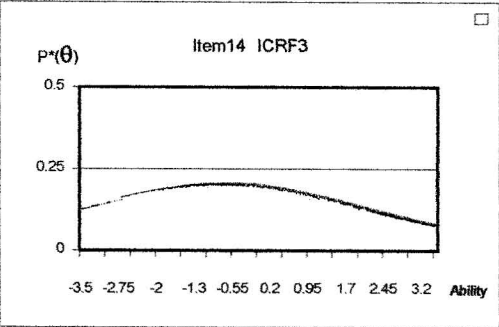
ข้อ 14 ลำดับขั้นที่ 1



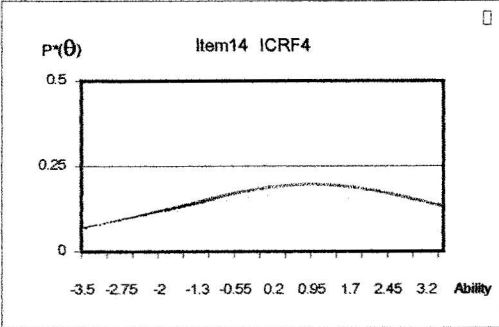
ลำดับขั้นที่ 2



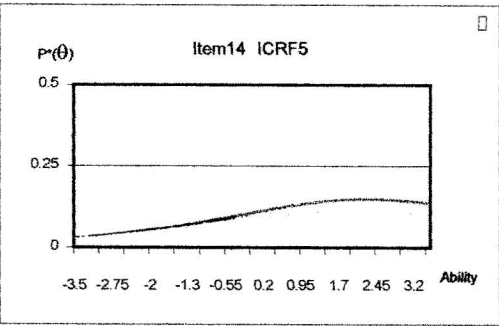
ลำดับชั้นที่ 3



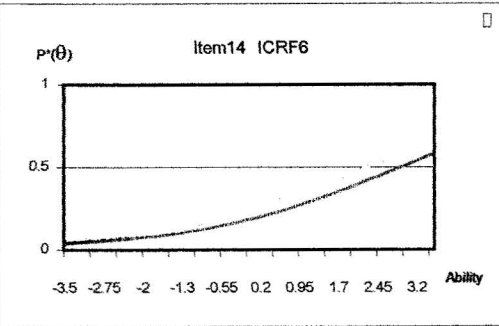
ลำดับชั้นที่ 4



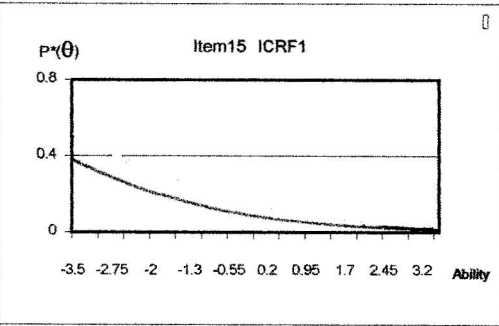
ลำดับชั้นที่ 5



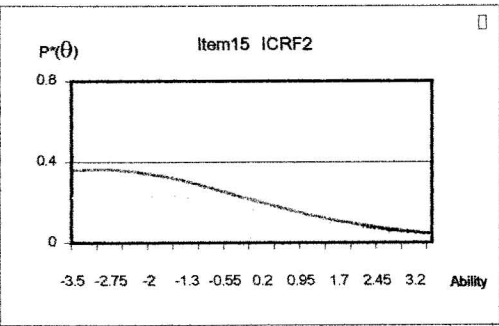
ลำดับชั้นที่ 6



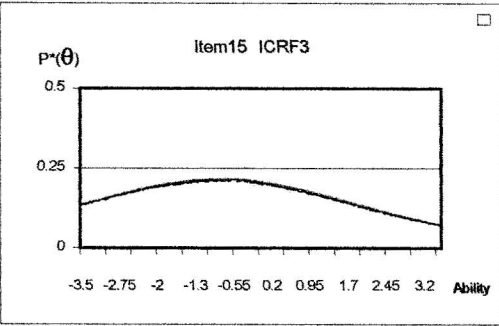
ข้อ 15 ลำดับชั้นที่ 1



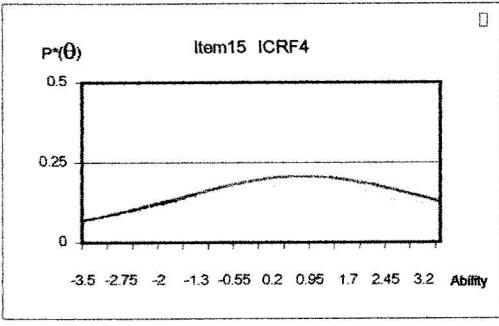
ลำดับชั้นที่ 2



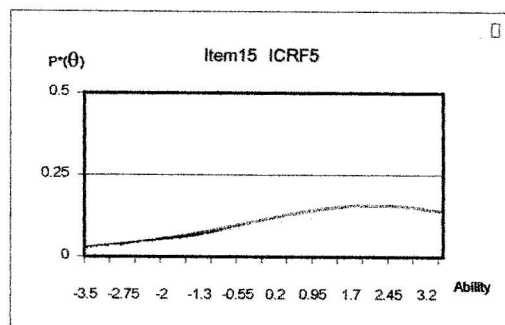
ลำดับชั้นที่ 3



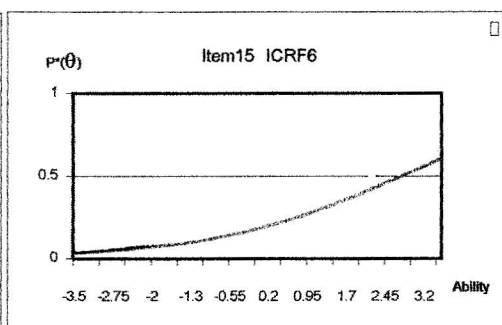
ลำดับชั้นที่ 4



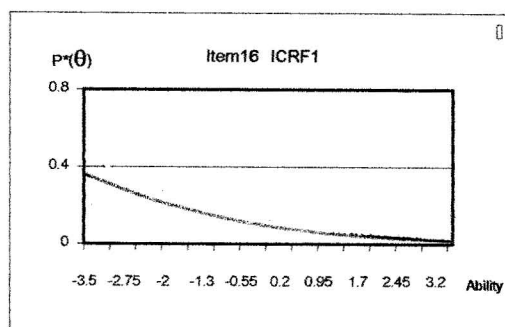
ลำดับขั้นที่ 5



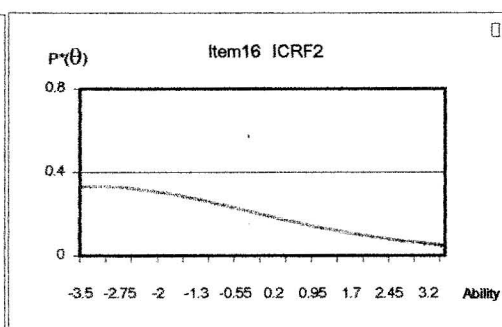
ลำดับขั้นที่ 6



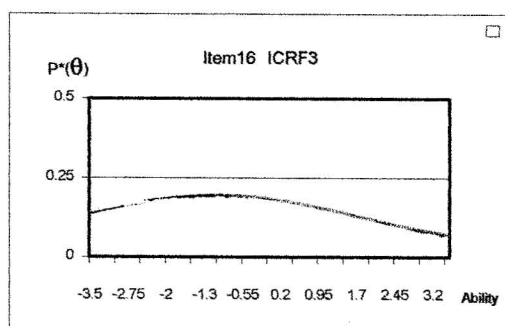
ข้อ 16 ลำดับขั้นที่ 1



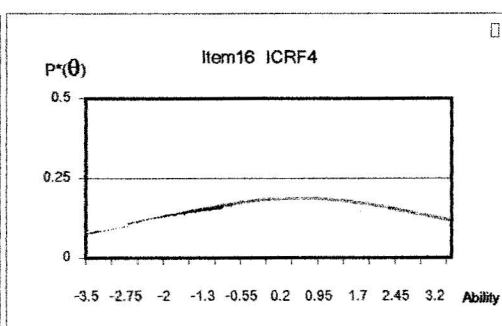
ลำดับขั้นที่ 2



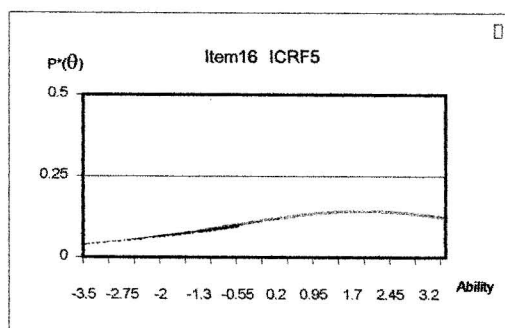
ลำดับขั้นที่ 3



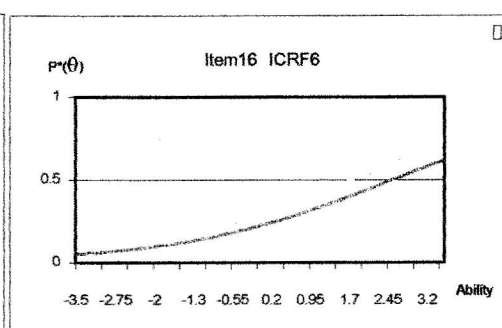
ลำดับขั้นที่ 4



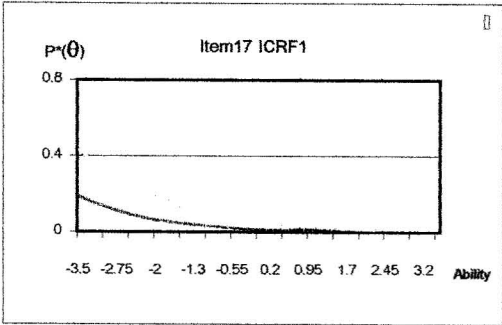
ลำดับขั้นที่ 5



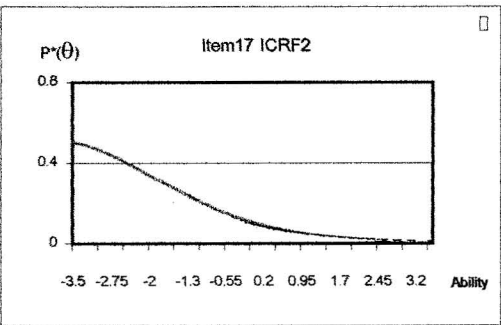
ลำดับขั้นที่ 6



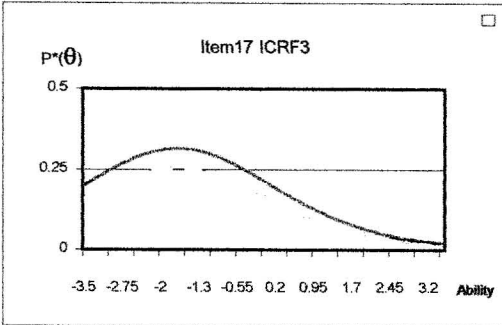
ข้อ 17 ลำดับขั้นที่ 1



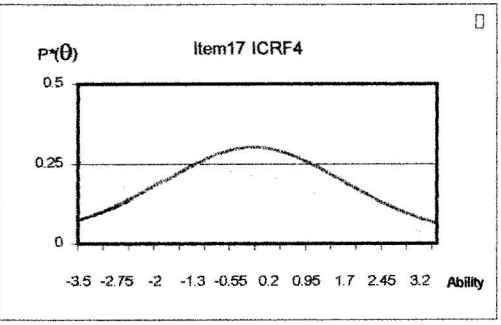
ลำดับขั้นที่ 2



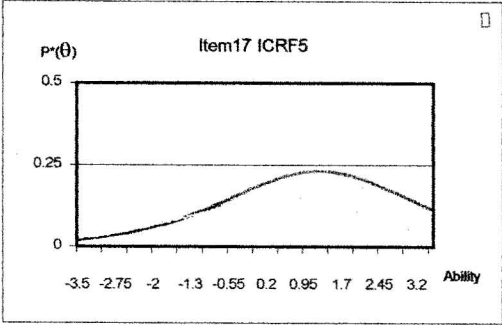
ลำดับขั้นที่ 3



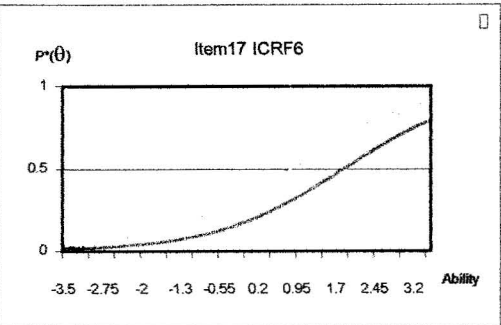
ลำดับขั้นที่ 4



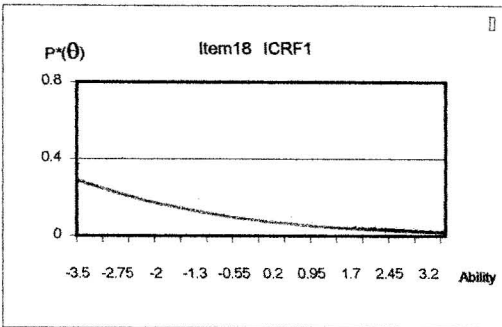
ลำดับขั้นที่ 5



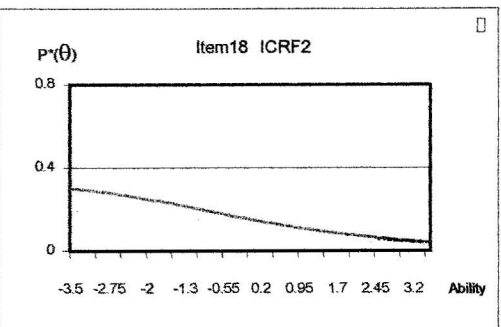
ลำดับขั้นที่ 6



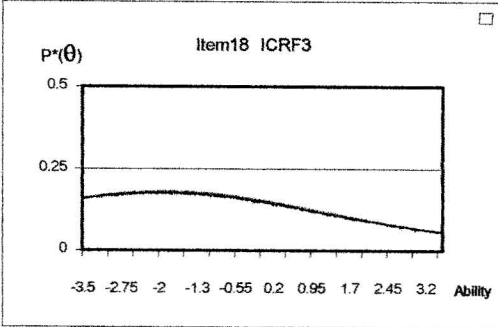
ข้อ 18 ลำดับขั้นที่ 1



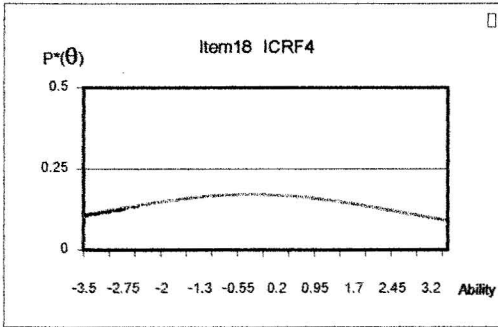
ลำดับขั้นที่ 2



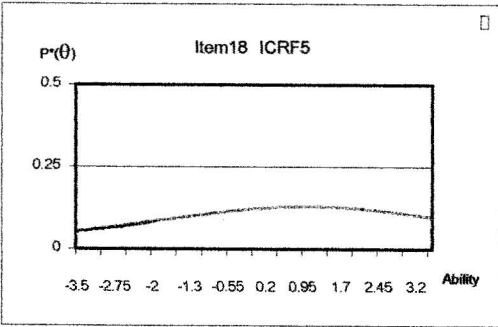
ลำดับชั้นที่ 3



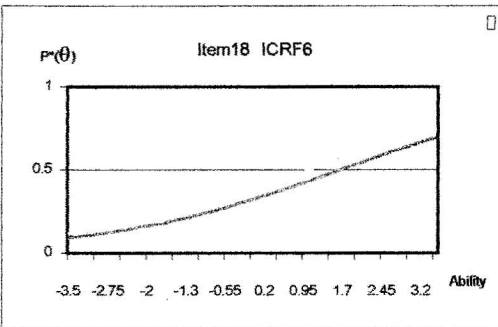
ลำดับชั้นที่ 4



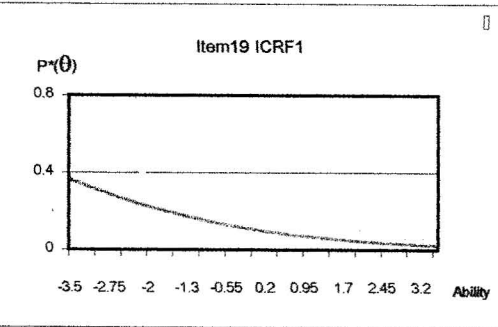
ลำดับชั้นที่ 5



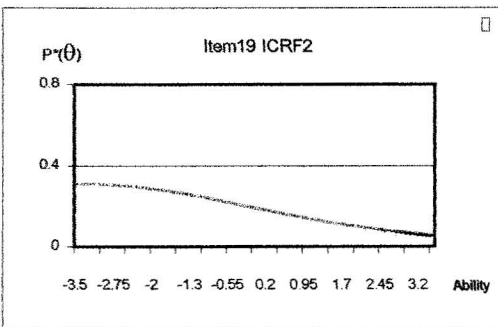
ลำดับชั้นที่ 6



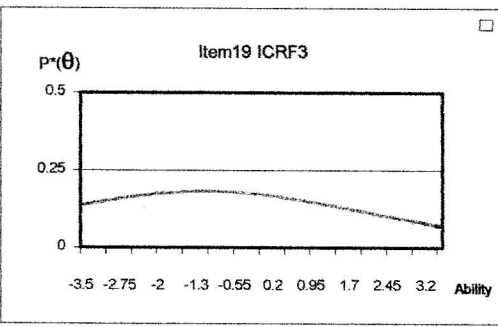
ข้อ 19 ลำดับชั้นที่ 1



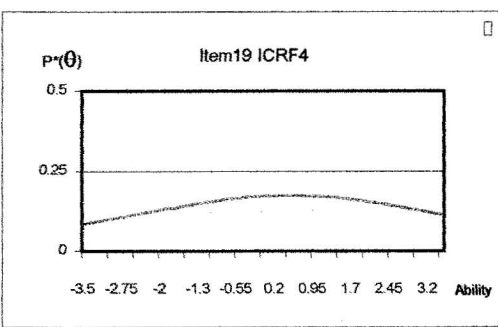
ลำดับชั้นที่ 2



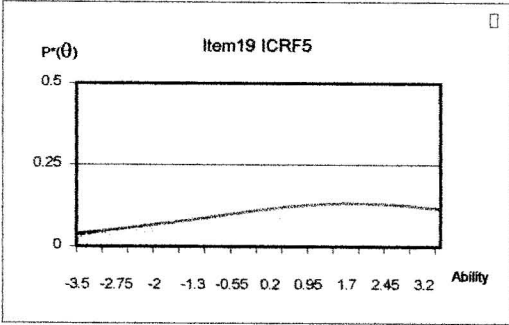
ลำดับชั้นที่ 3



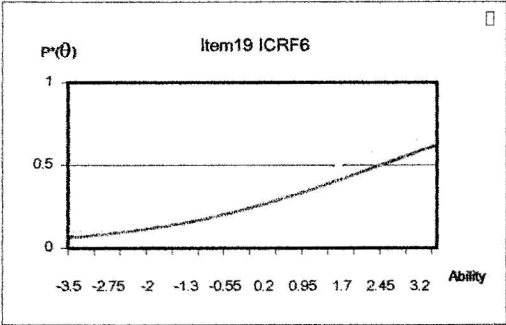
ลำดับชั้นที่ 4



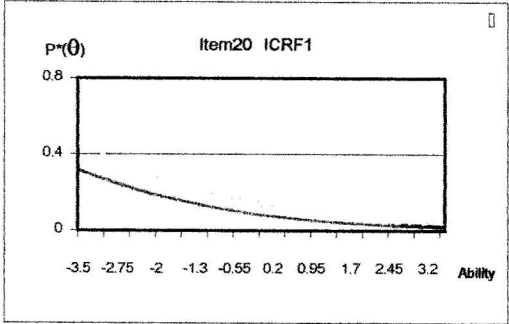
ลำดับชั้นที่ 5



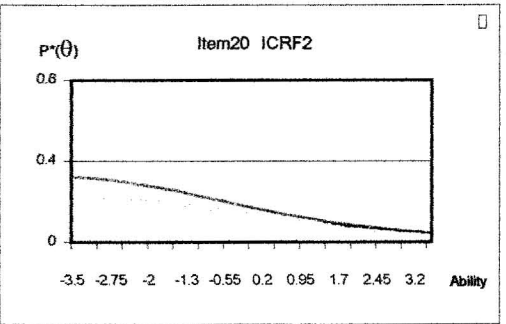
ลำดับชั้นที่ 6



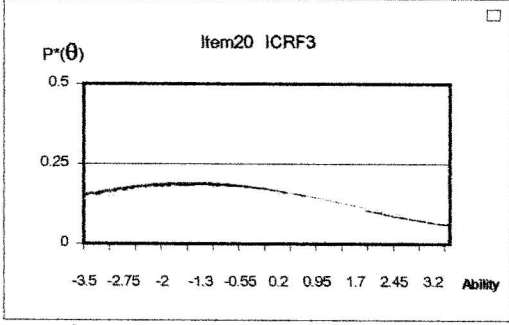
ข้อ 20 ลำดับชั้นที่ 1



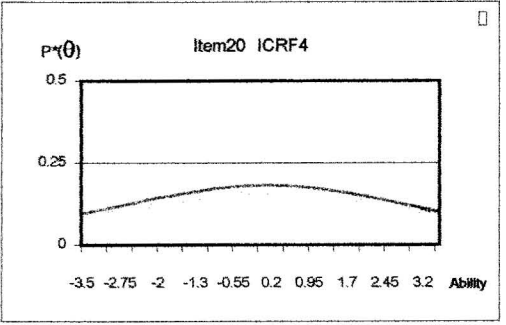
ลำดับชั้นที่ 2



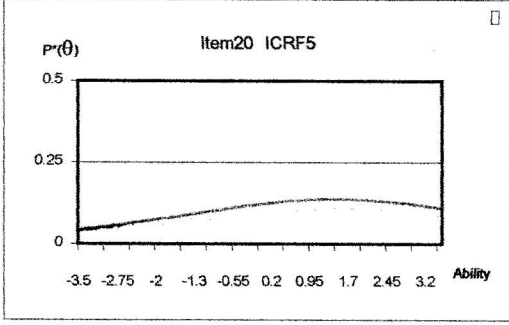
ลำดับชั้นที่ 3



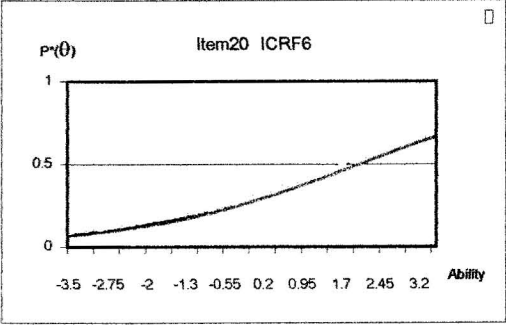
ลำดับชั้นที่ 4



ลำดับชั้นที่ 5



ลำดับชั้นที่ 6



ภาคผนวก ฐ

ฟังก์ชันผลการตอบข้อสอบในขอบเขตลำดับชั้น

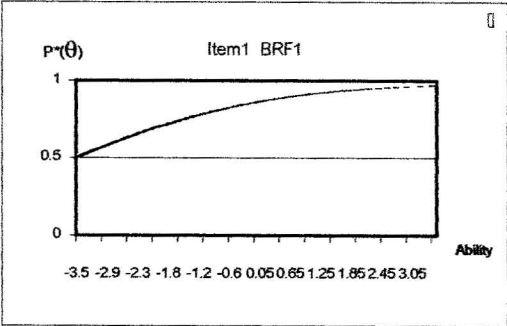
[Boundary Response Function : BRF]

- แบบทดสอบขนาด 40 ข้อ
- แบบทดสอบขนาด 30 ข้อ
- แบบทดสอบขนาด 20 ข้อ

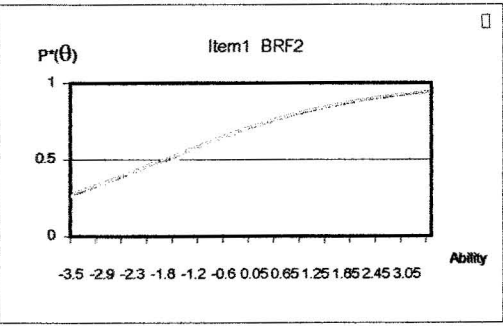
ฟังก์ชันผลการตอบข้อสอบในขอบเขตลำดับชั้น (BRF) ที่วิเคราะห์ด้วยวิธี GPCM ในแบบทดสอบขนาด 40 ข้อ
เมื่อกลุ่มตัวอย่างเพศชายต่อเพศหญิง เท่ากับ 1,000 คน ต่อ 1,000 คน

ชาย หญิง

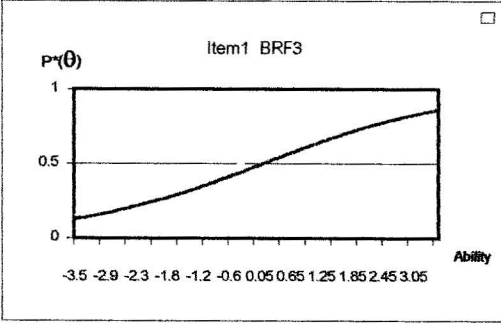
ข้อ 1 ขอบเขตลำดับชั้นที่ 1



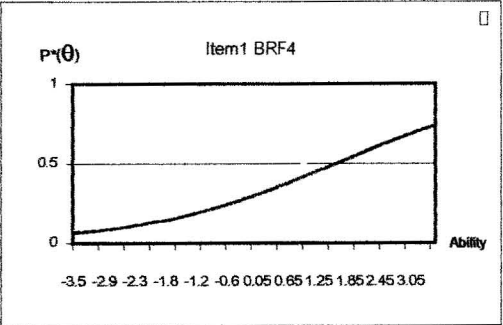
ขอบเขตลำดับชั้นที่ 2



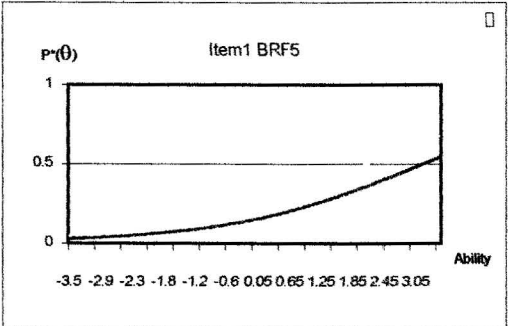
ขอบเขตลำดับชั้นที่ 3



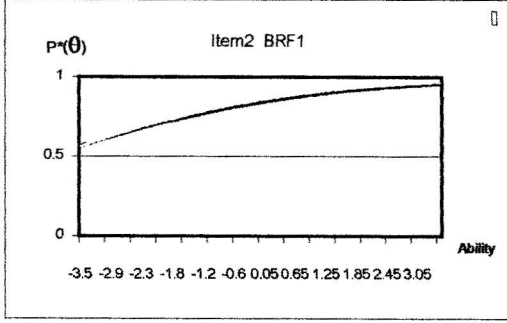
ขอบเขตลำดับชั้นที่ 4



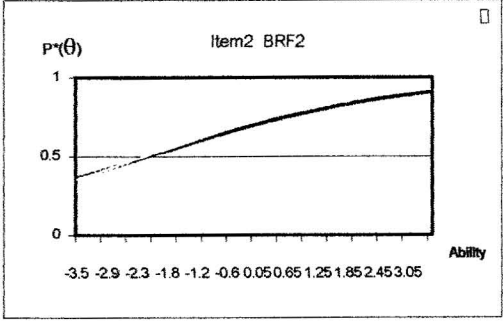
ขอบเขตลำดับชั้นที่ 5



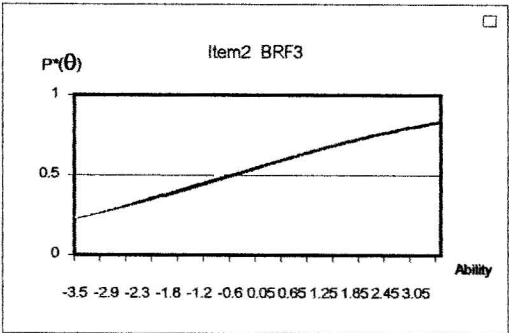
ข้อ 2 ขอบเขตลำดับชั้นที่ 1



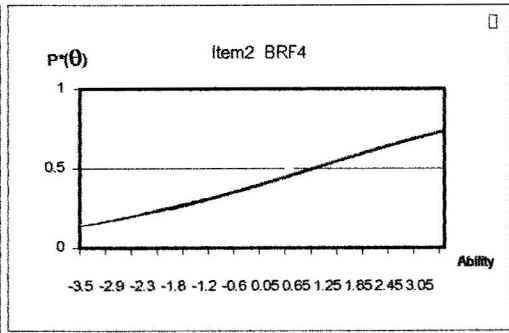
ขอบเขตลำดับชั้นที่ 2



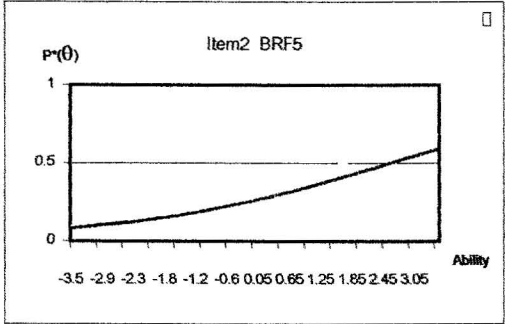
ข้อ 3 ขอบเขตลำดับขั้นที่ 3



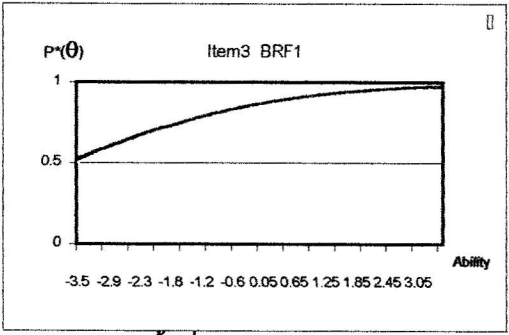
ข้อ 3 ขอบเขตลำดับขั้นที่ 4



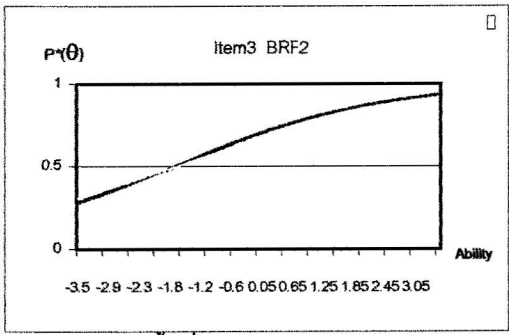
ข้อ 3 ขอบเขตลำดับขั้นที่ 5



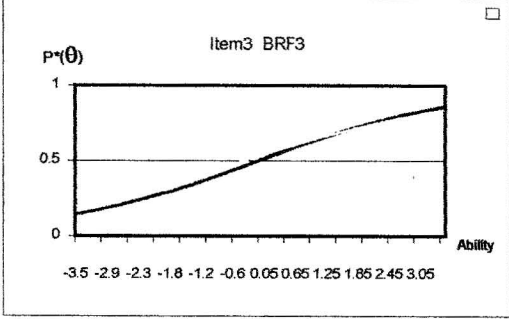
ข้อ 3 ขอบเขตลำดับขั้นที่ 1



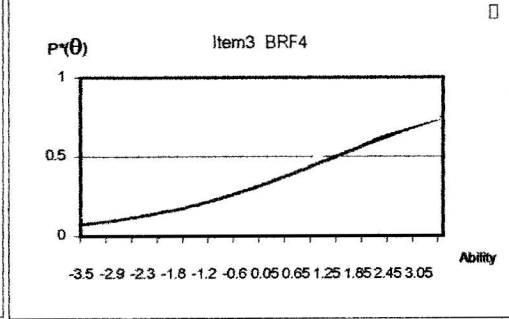
ข้อ 3 ขอบเขตลำดับขั้นที่ 2



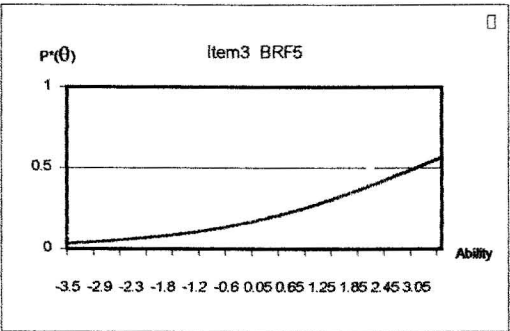
ข้อ 3 ขอบเขตลำดับขั้นที่ 3



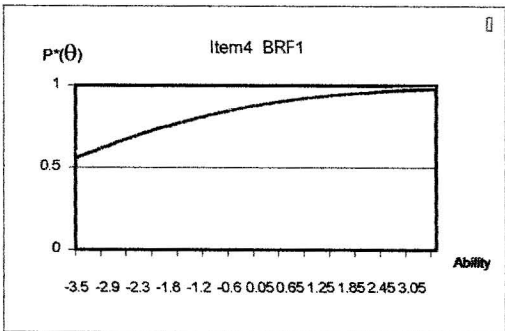
ข้อ 3 ขอบเขตลำดับขั้นที่ 4



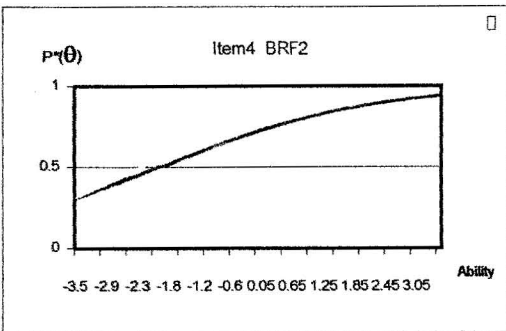
ขอบเขตลำดับขั้นที่ 5



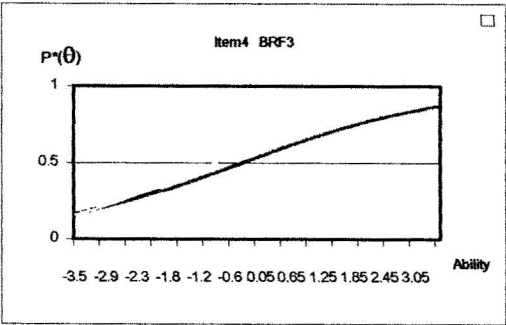
ข้อ 4 ขอบเขตลำดับขั้นที่ 1



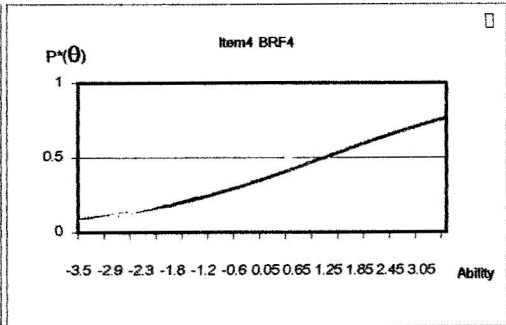
ขอบเขตลำดับขั้นที่ 2



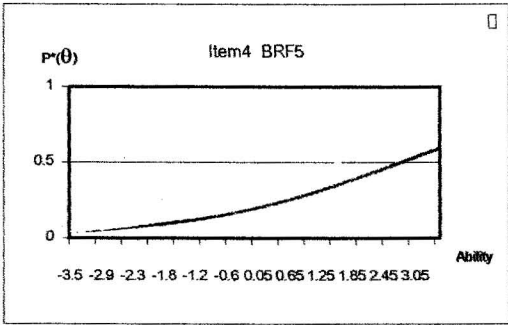
ขอบเขตลำดับขั้นที่ 3



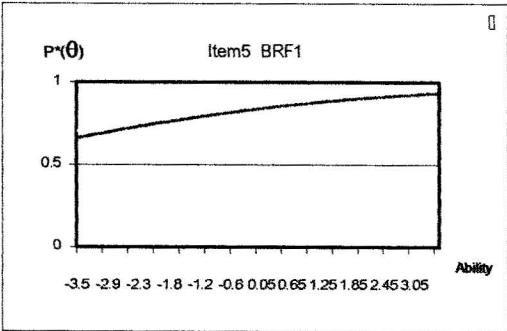
ขอบเขตลำดับขั้นที่ 4



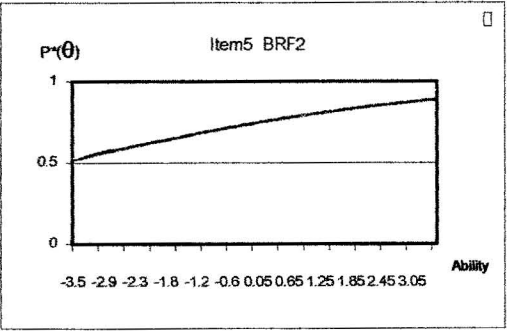
ขอบเขตลำดับขั้นที่ 5



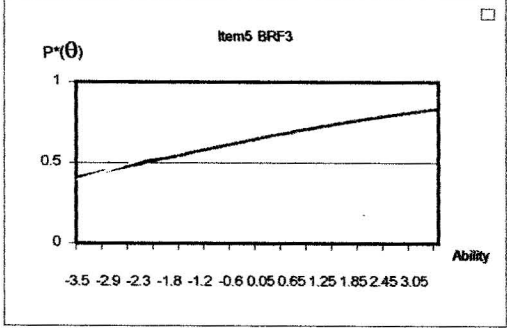
ข้อ 5 ขอบเขตลำดับขั้นที่ 1



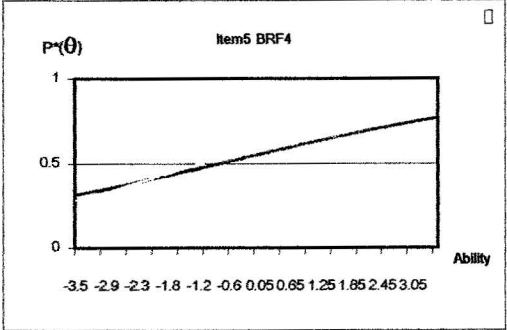
ข้อ 5 ขอบเขตลำดับขั้นที่ 2



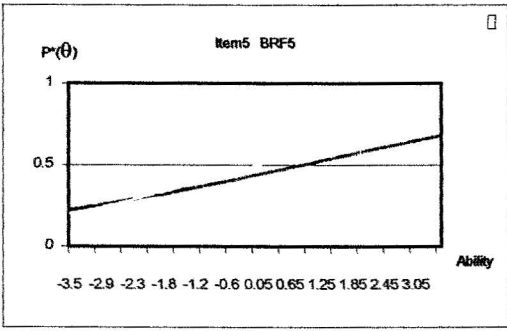
ข้อ 5 ขอบเขตลำดับขั้นที่ 3



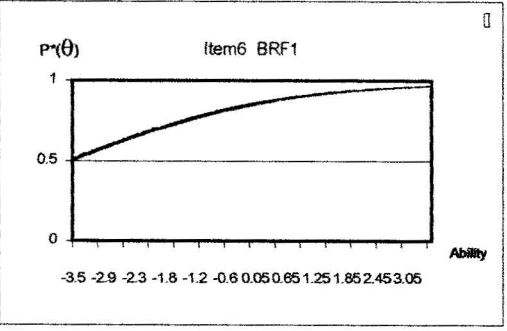
ข้อ 5 ขอบเขตลำดับขั้นที่ 4



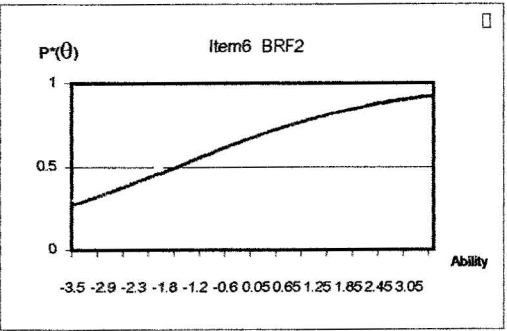
ข้อ 5 ขอบเขตลำดับขั้นที่ 5



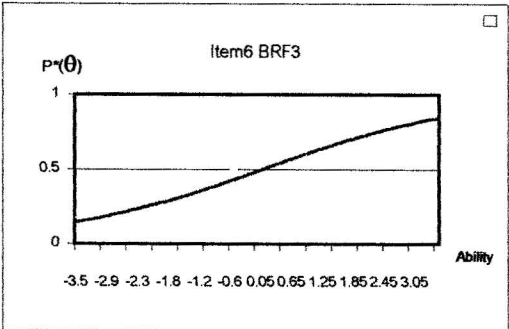
ข้อ 6 ขอบเขตลำดับขั้นที่ 1



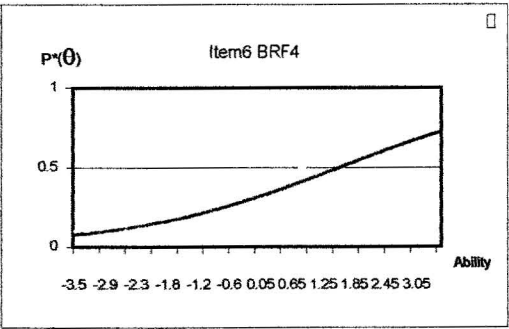
ข้อ 6 ขอบเขตลำดับขั้นที่ 2



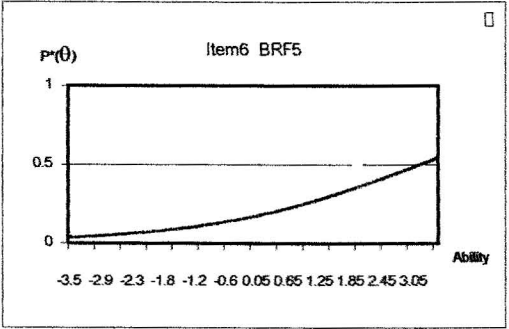
ข้อ 7 ขอบเขตลำดับขั้นที่ 3



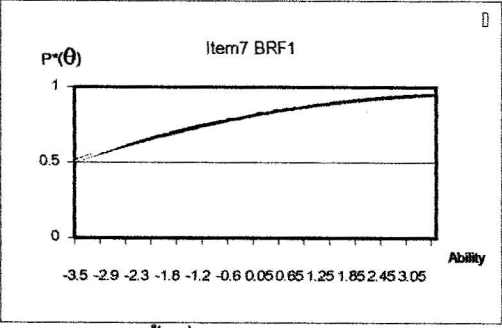
ข้อ 7 ขอบเขตลำดับขั้นที่ 4



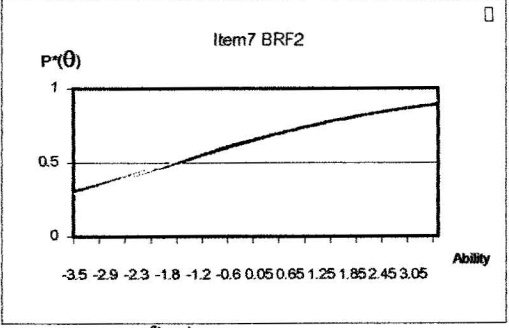
ข้อ 7 ขอบเขตลำดับขั้นที่ 5



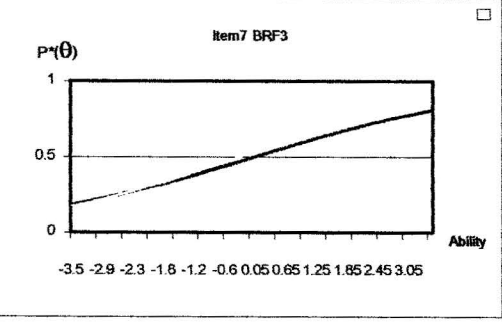
ข้อ 7 ขอบเขตลำดับขั้นที่ 1



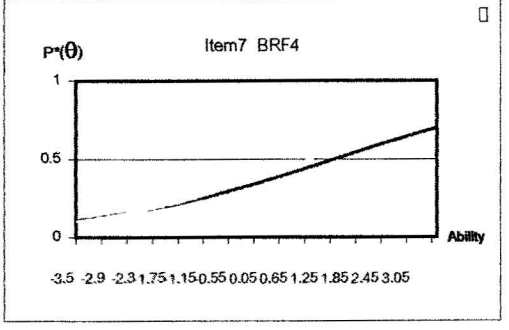
ข้อ 7 ขอบเขตลำดับขั้นที่ 2



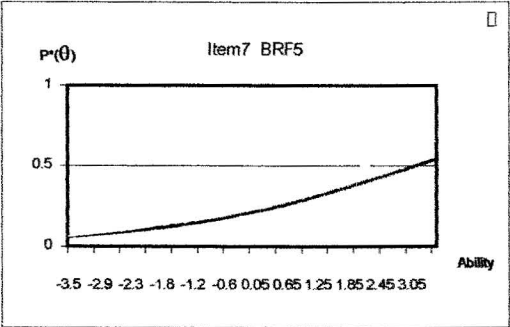
ข้อ 7 ขอบเขตลำดับขั้นที่ 3



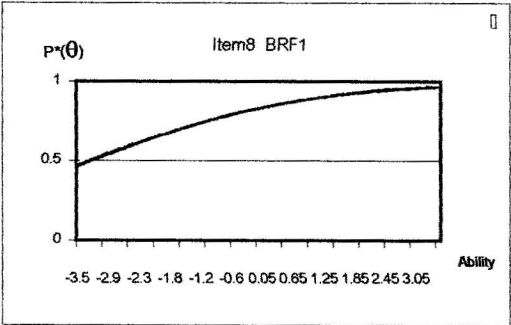
ข้อ 7 ขอบเขตลำดับขั้นที่ 4



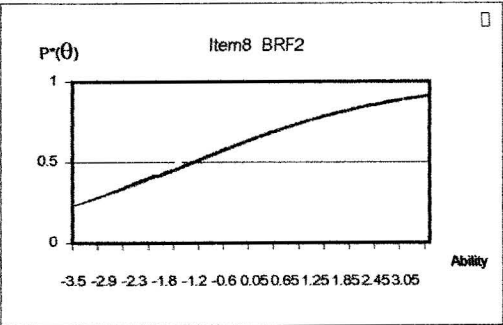
ขอบเขตลำดับขั้นที่ 5



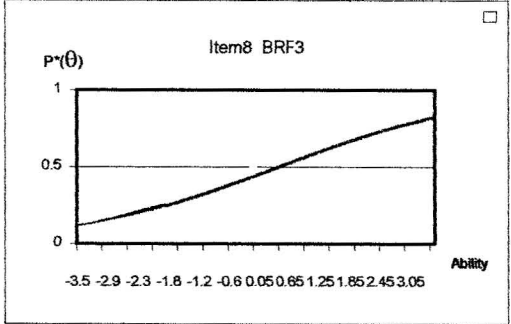
ข้อ 8 ขอบเขตลำดับขั้นที่ 1



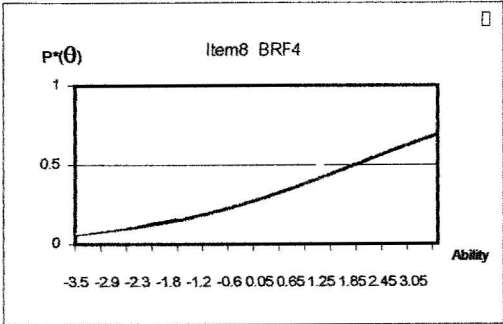
ขอบเขตลำดับขั้นที่ 2



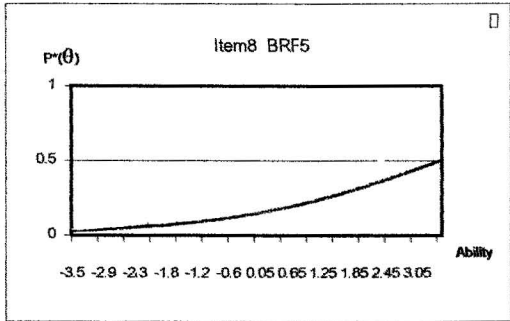
ขอบเขตลำดับขั้นที่ 3



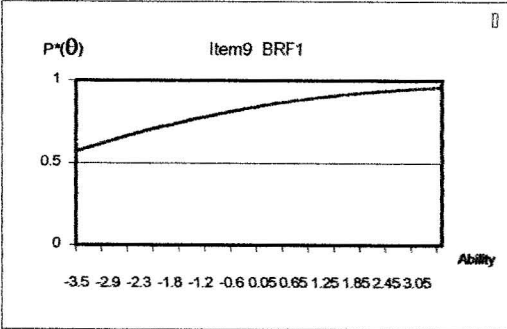
ขอบเขตลำดับขั้นที่ 4



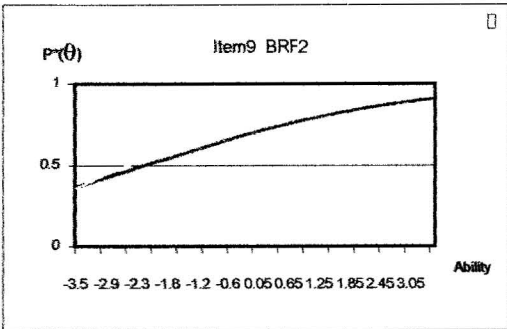
ขอบเขตลำดับขั้นที่ 5



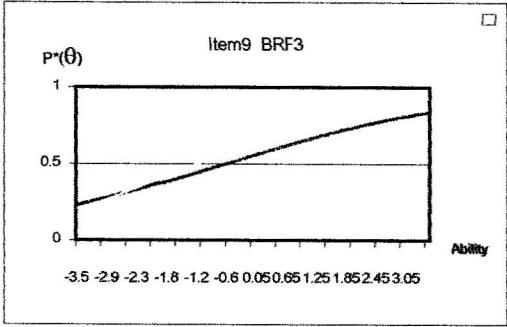
ข้อ 9 ขอบเขตลำดับขั้นที่ 1



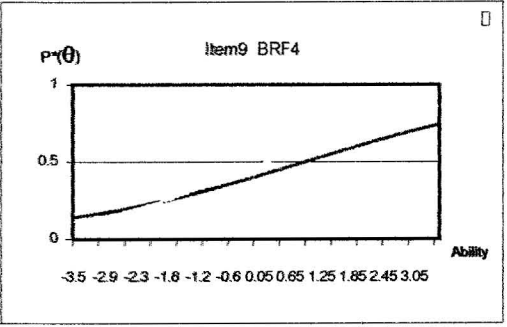
ขอบเขตลำดับขั้นที่ 2



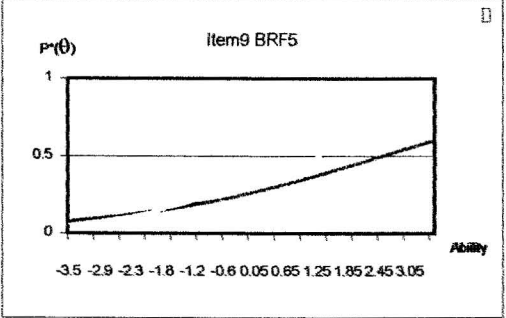
ขอบเขตลำดับขั้นที่ 3



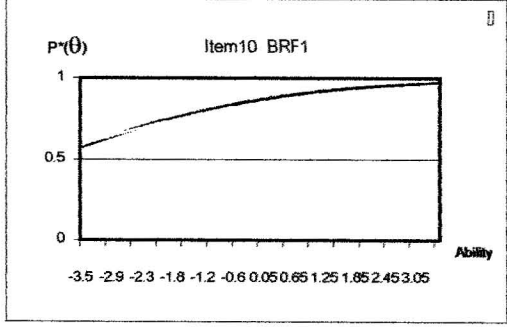
ขอบเขตลำดับขั้นที่ 4



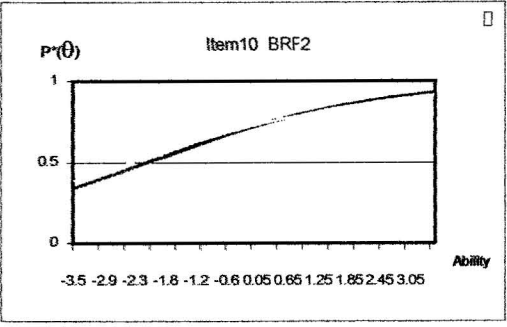
ขอบเขตลำดับขั้นที่ 5



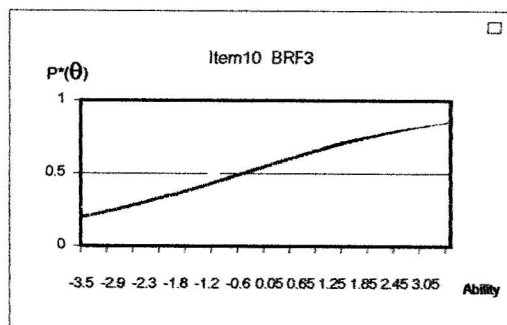
ข้อ 10 ขอบเขตลำดับขั้นที่ 1



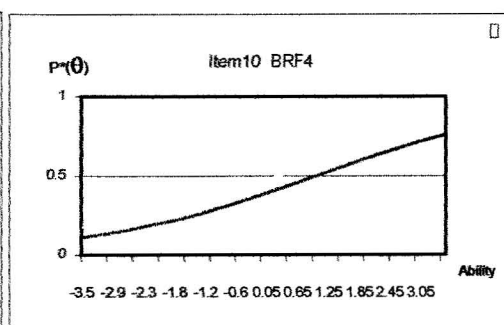
ขอบเขตลำดับขั้นที่ 2



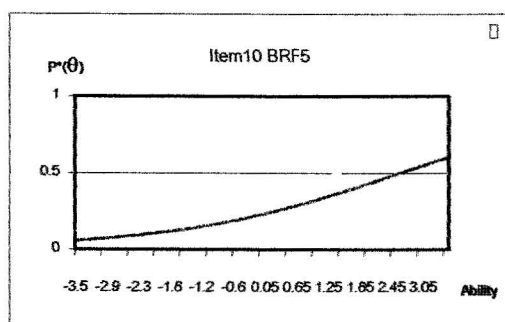
ขอบเขตลำดับขั้นที่ 3



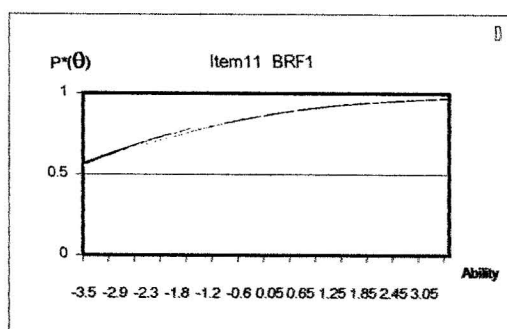
ขอบเขตลำดับขั้นที่ 4



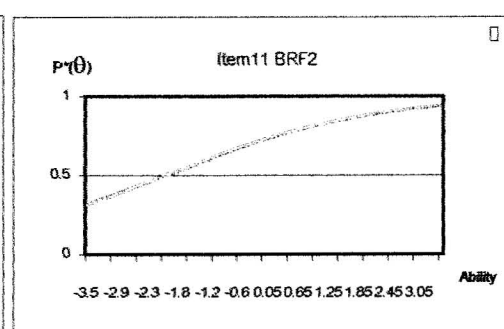
ขอบเขตลำดับขั้นที่ 5



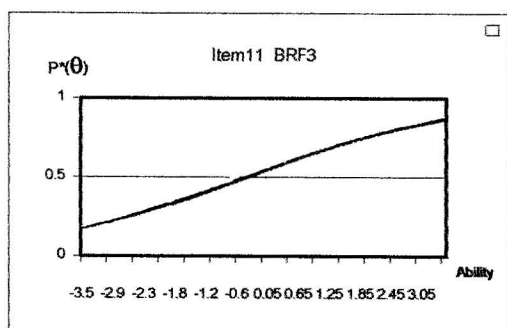
ข้อ 11 ขอบเขตลำดับขั้นที่ 1



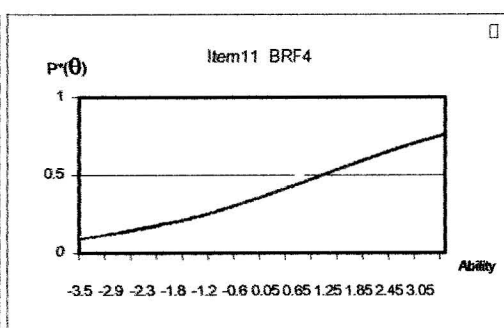
ขอบเขตลำดับขั้นที่ 2



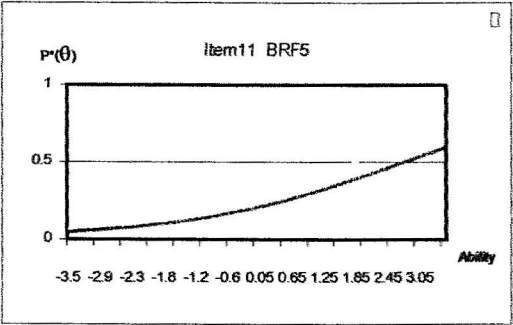
ขอบเขตลำดับขั้นที่ 3



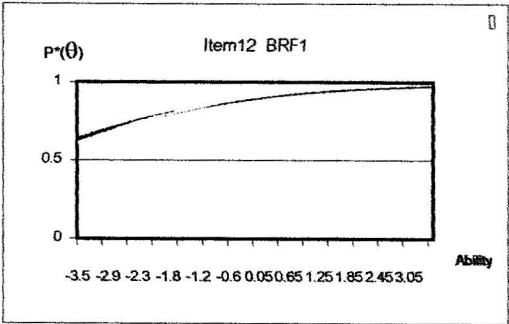
ขอบเขตลำดับขั้นที่ 4



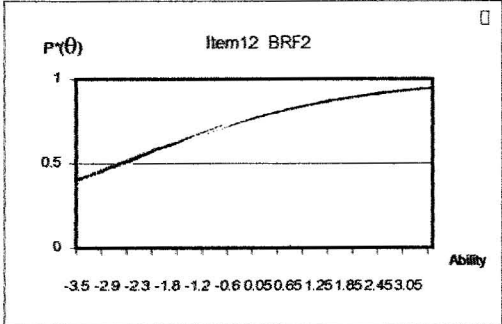
ขอบเขตลำดับขั้นที่ 5



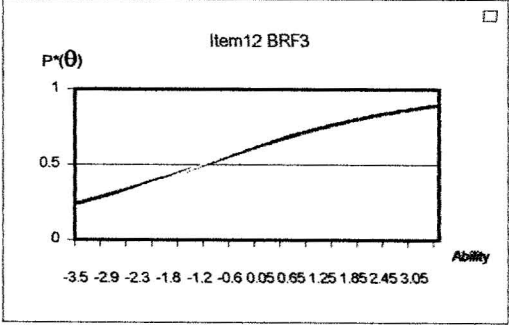
ข้อ 12 ขอบเขตลำดับขั้นที่ 1



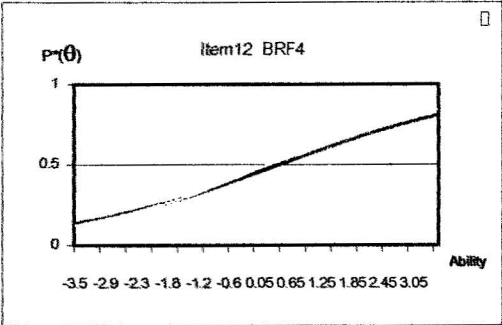
ขอบเขตลำดับขั้นที่ 2



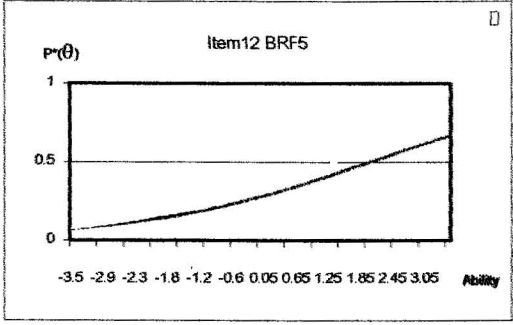
ขอบเขตลำดับขั้นที่ 3



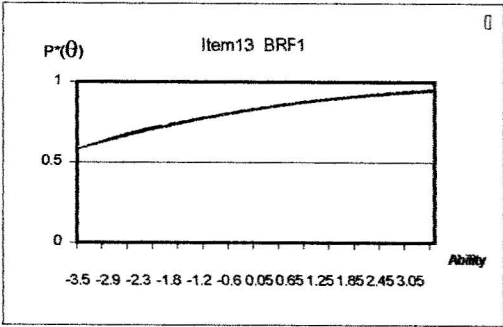
ขอบเขตลำดับขั้นที่ 4



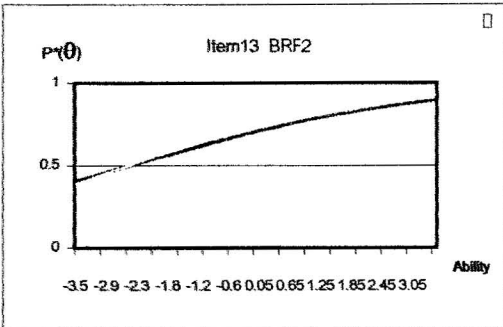
ขอบเขตลำดับขั้นที่ 5



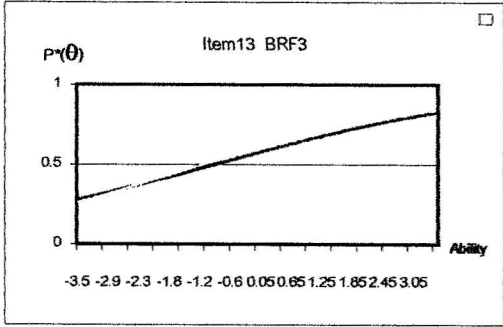
ข้อ 13 ขอบเขตลำดับขั้นที่ 1



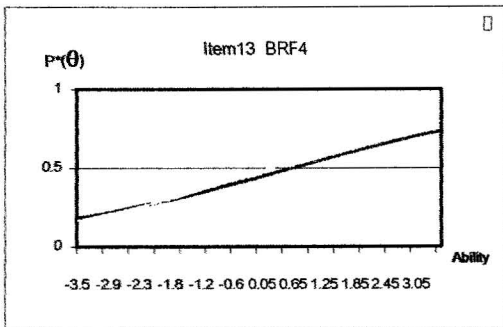
ข้อ 13 ขอบเขตลำดับขั้นที่ 2



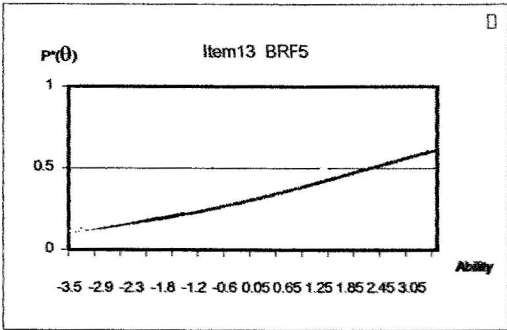
ข้อ 13 ขอบเขตลำดับขั้นที่ 3



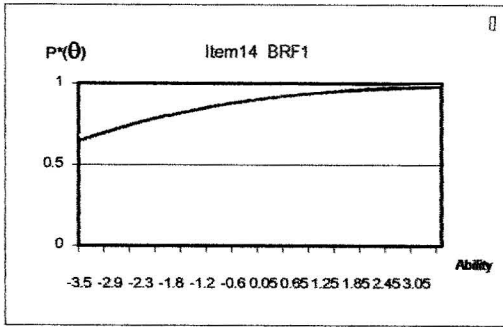
ข้อ 13 ขอบเขตลำดับขั้นที่ 4



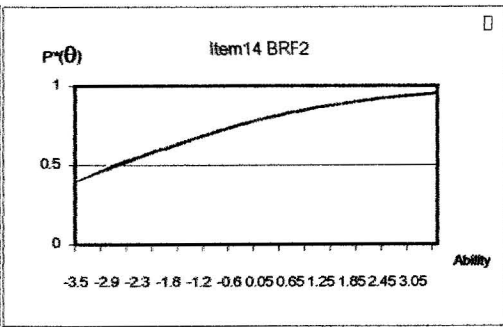
ข้อ 13 ขอบเขตลำดับขั้นที่ 5



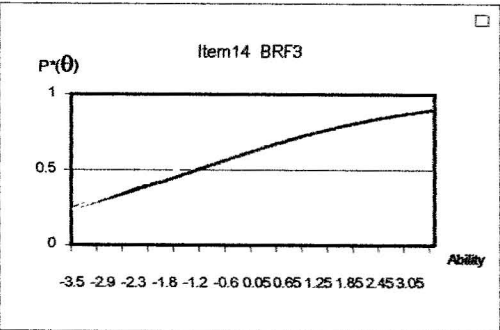
ข้อ 14 ขอบเขตลำดับขั้นที่ 1



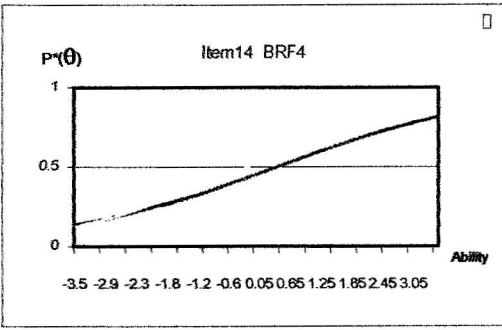
ข้อ 14 ขอบเขตลำดับขั้นที่ 2



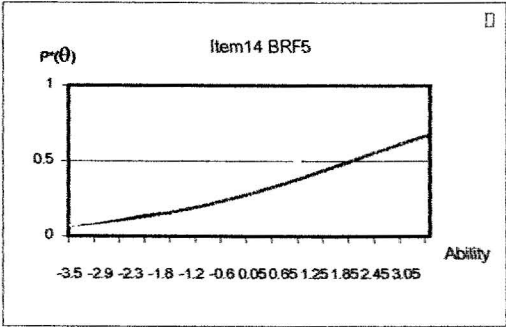
ขอบเขตลำดับขั้นที่ 3



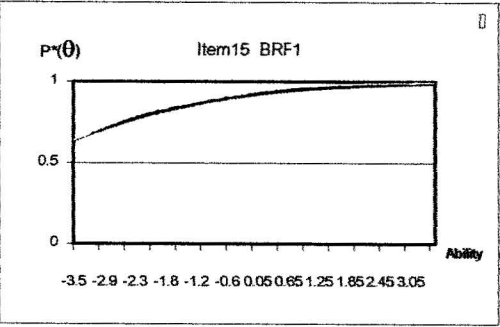
ขอบเขตลำดับขั้นที่ 4



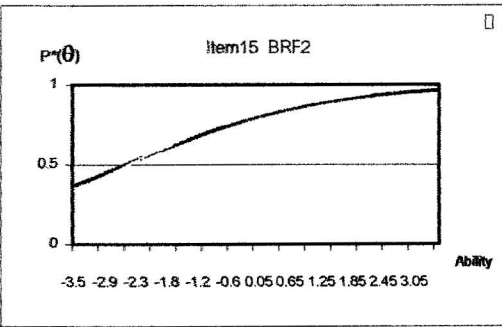
ขอบเขตลำดับขั้นที่ 5



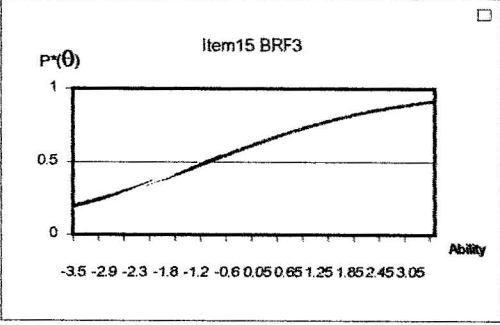
ข้อ 15 ขอบเขตลำดับขั้นที่ 1



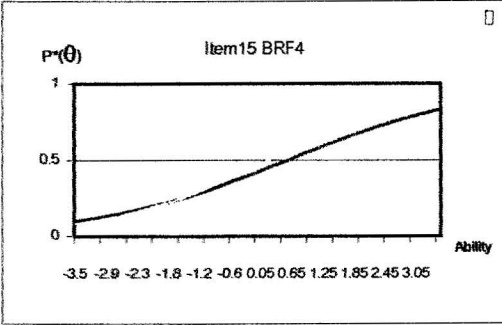
ขอบเขตลำดับขั้นที่ 2



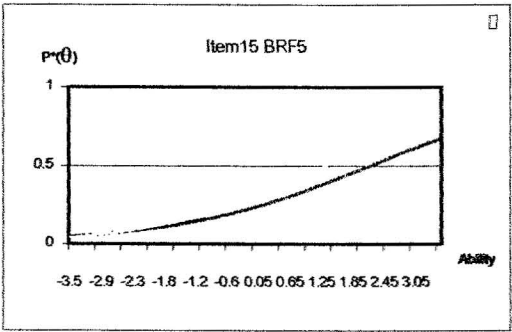
ขอบเขตลำดับขั้นที่ 3



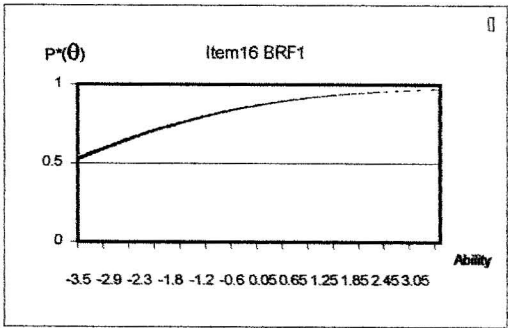
ขอบเขตลำดับขั้นที่ 4



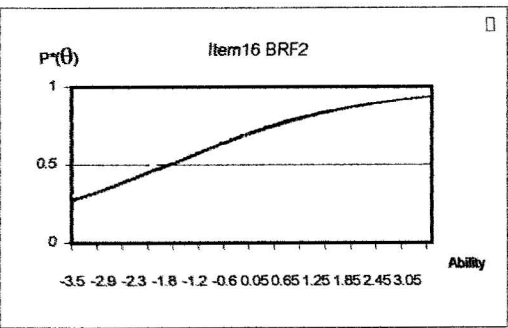
ขอบเขตลำดับขั้นที่ 5



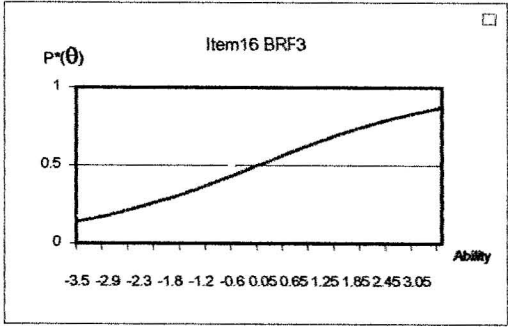
ข้อ 16 ขอบเขตลำดับขั้นที่ 1



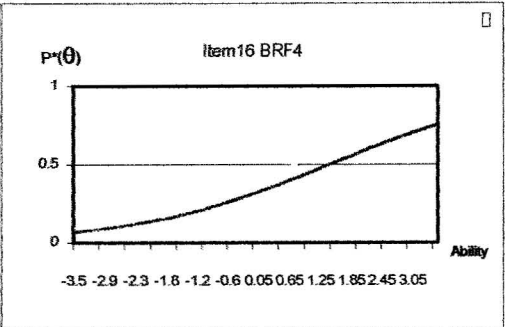
ขอบเขตลำดับขั้นที่ 2



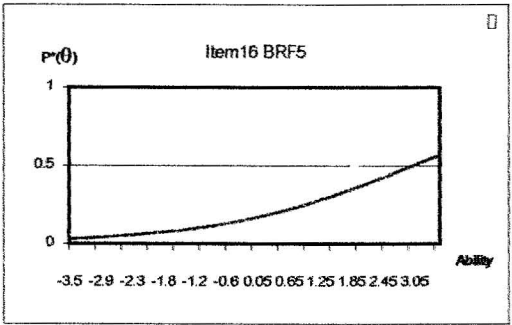
ขอบเขตลำดับขั้นที่ 3



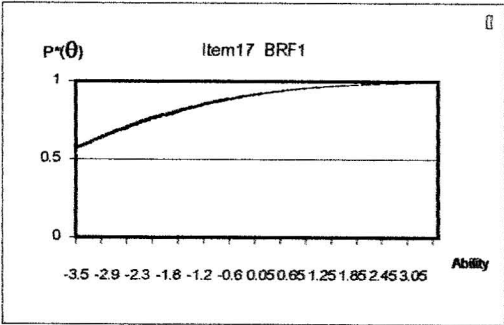
ขอบเขตลำดับขั้นที่ 4



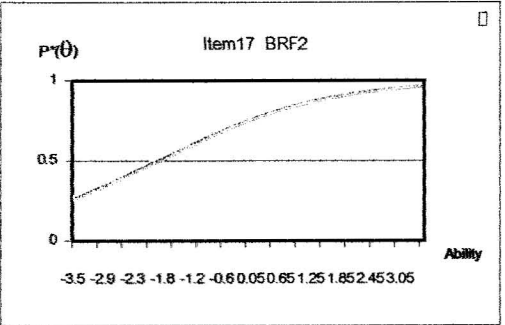
ขอบเขตลำดับขั้นที่ 5



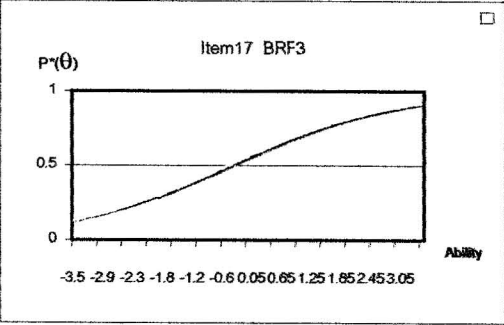
ข้อ 17 ขอบเขตลำดับขั้นที่ 1



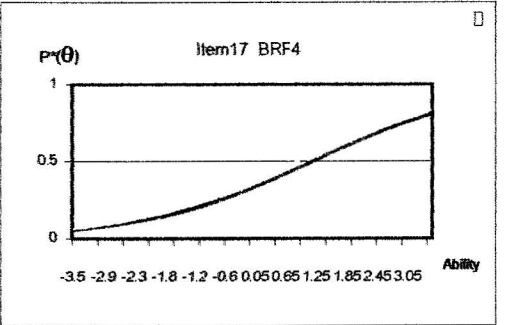
ข้อ 17 ขอบเขตลำดับขั้นที่ 2



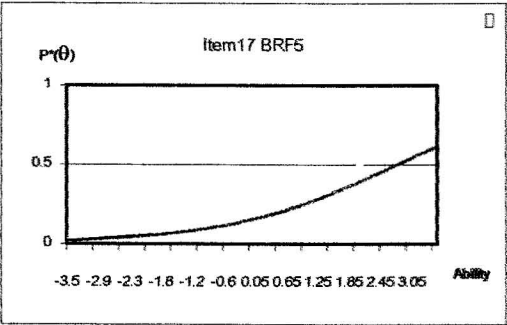
ข้อ 17 ขอบเขตลำดับขั้นที่ 3



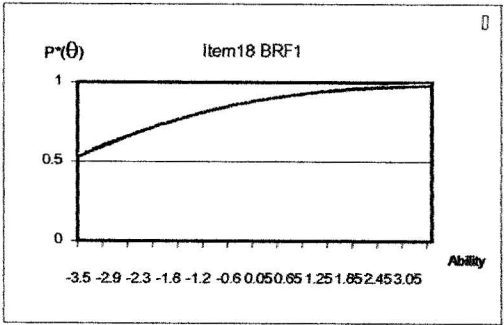
ข้อ 17 ขอบเขตลำดับขั้นที่ 4



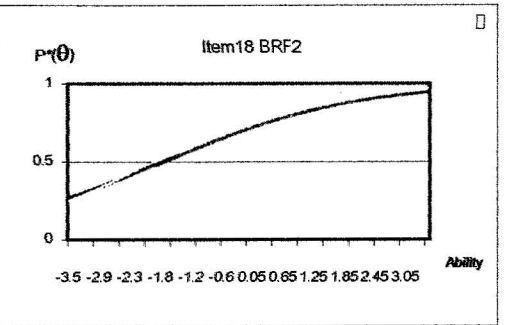
ข้อ 17 ขอบเขตลำดับขั้นที่ 5



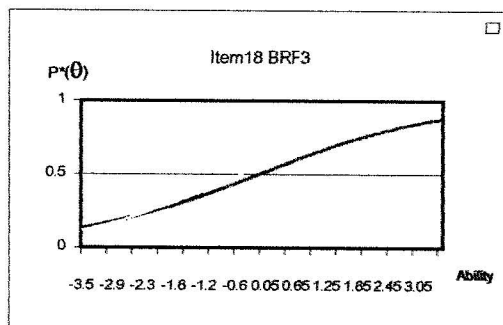
ข้อ 18 ขอบเขตลำดับขั้นที่ 1



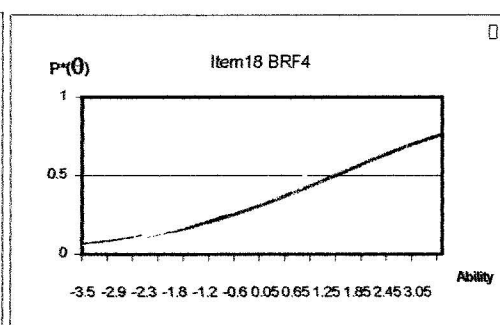
ข้อ 18 ขอบเขตลำดับขั้นที่ 2



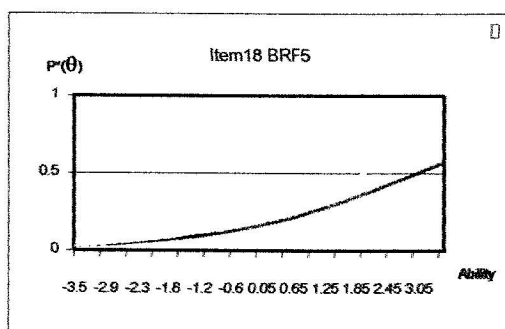
ขอบเขตลำดับขั้นที่ 3



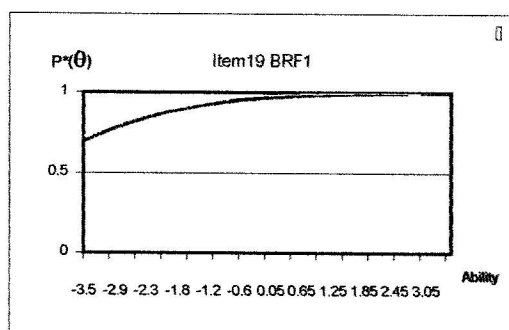
ขอบเขตลำดับขั้นที่ 4



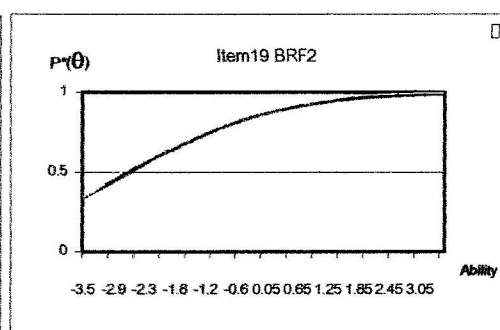
ขอบเขตลำดับขั้นที่ 5



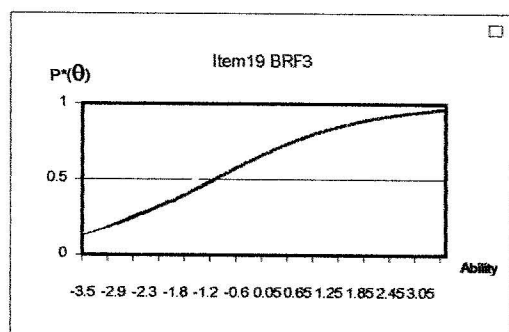
ข้อ 19 ขอบเขตลำดับขั้นที่ 1



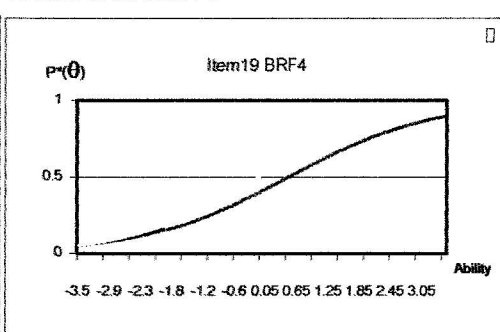
ขอบเขตลำดับขั้นที่ 2



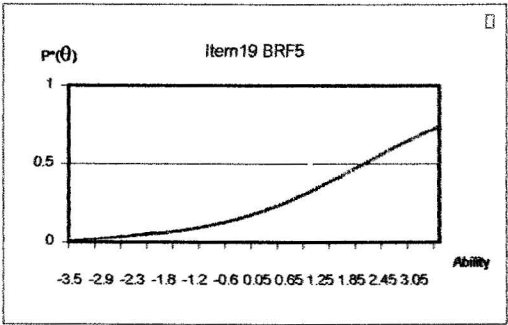
ขอบเขตลำดับขั้นที่ 3



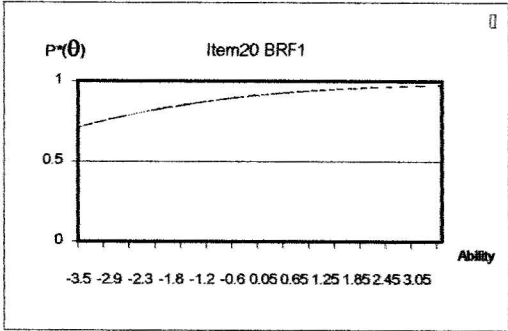
ขอบเขตลำดับขั้นที่ 4



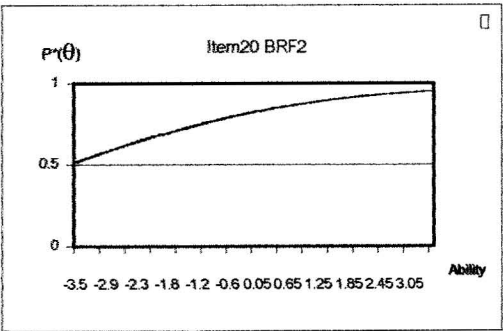
ขอบเขตลำดับขั้นที่ 5



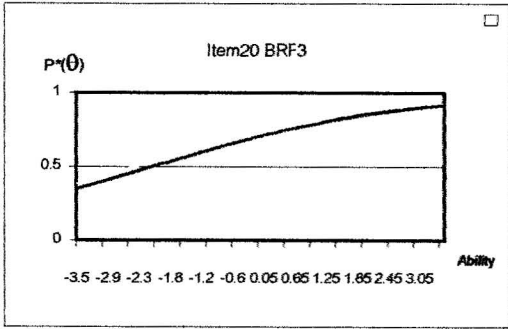
ข้อ 20 ขอบเขตลำดับขั้นที่ 1



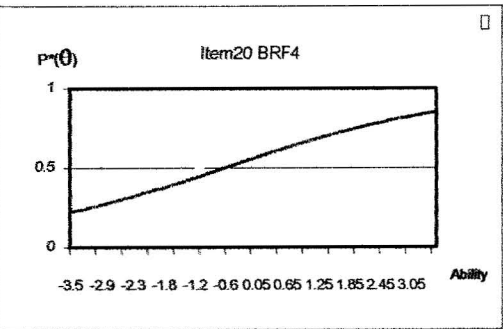
ขอบเขตลำดับขั้นที่ 2



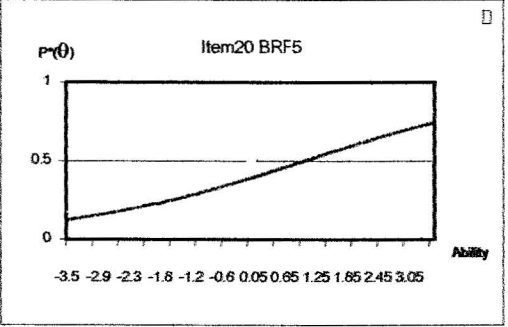
ขอบเขตลำดับขั้นที่ 3



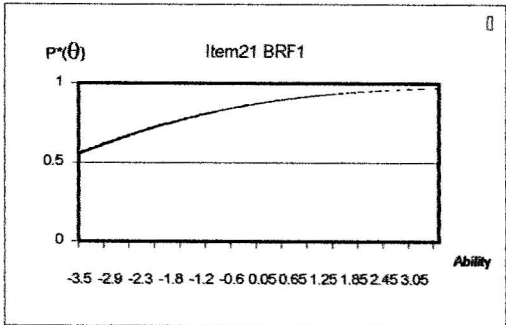
ขอบเขตลำดับขั้นที่ 4



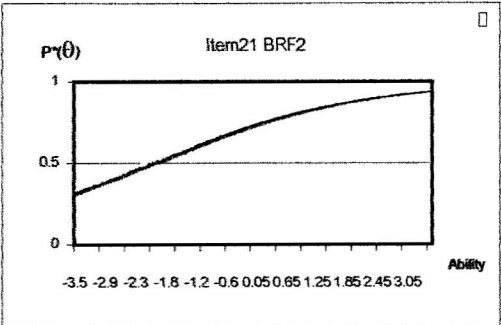
ขอบเขตลำดับขั้นที่ 5



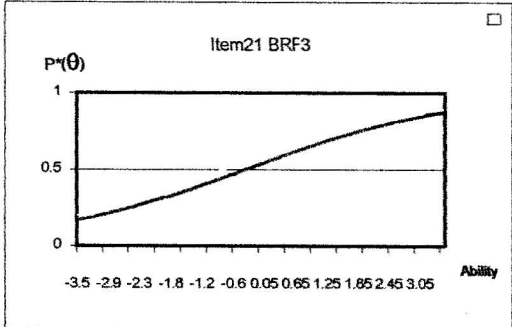
ข้อ 21 ขอบเขตลำดับขั้นที่ 1



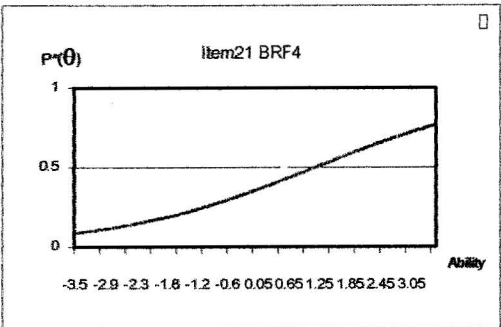
ข้อ 21 ขอบเขตลำดับขั้นที่ 2



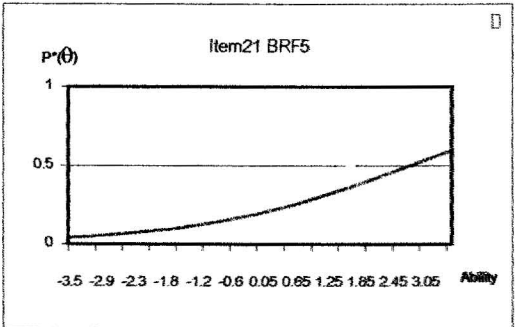
ข้อ 21 ขอบเขตลำดับขั้นที่ 3



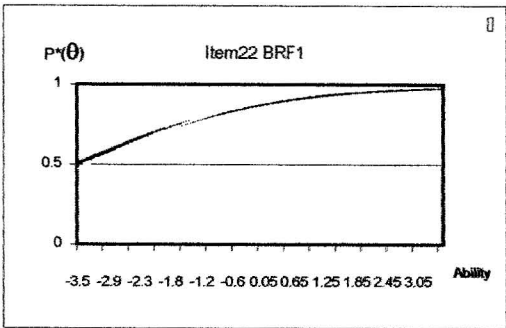
ข้อ 21 ขอบเขตลำดับขั้นที่ 4



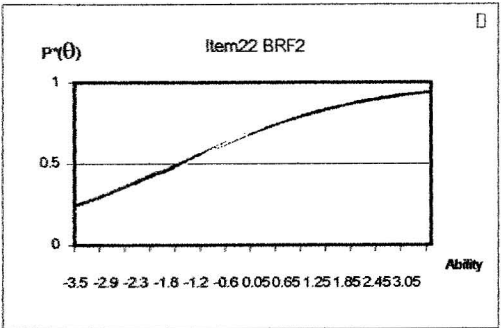
ข้อ 21 ขอบเขตลำดับขั้นที่ 5



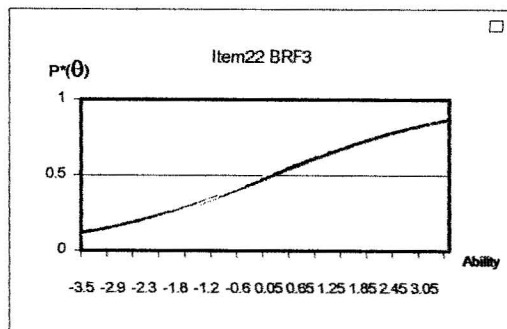
ข้อ 22 ขอบเขตลำดับขั้นที่ 1



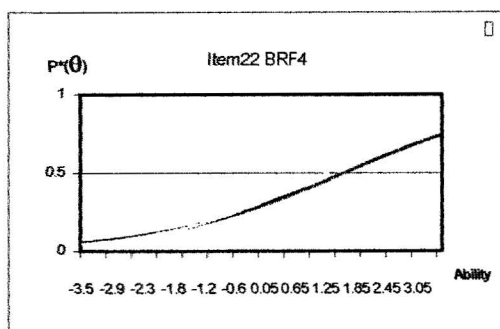
ข้อ 22 ขอบเขตลำดับขั้นที่ 2



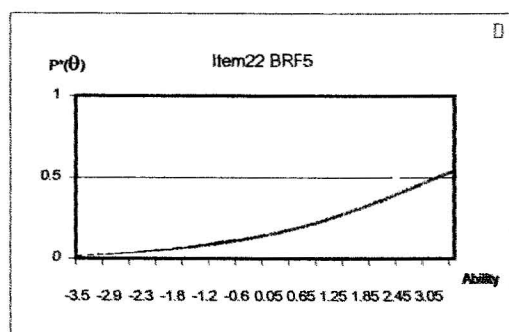
ขอบเขตลำดับขั้นที่ 3



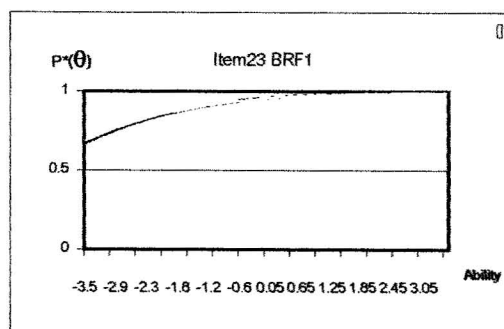
ขอบเขตลำดับขั้นที่ 4



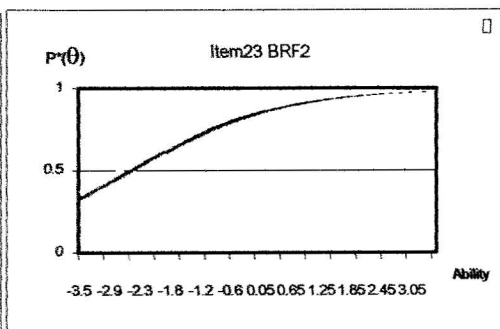
ขอบเขตลำดับขั้นที่ 5



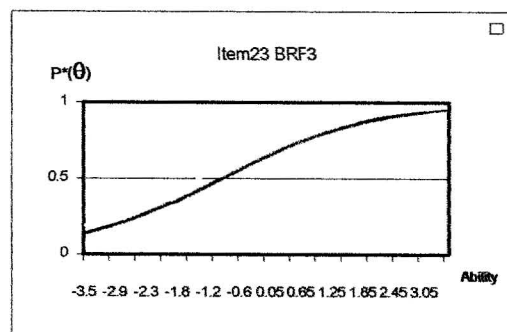
ข้อ 23 ขอบเขตลำดับขั้นที่ 1



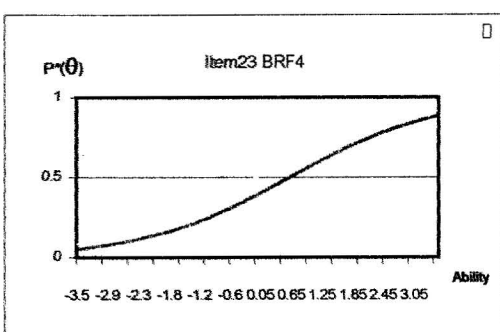
ขอบเขตลำดับขั้นที่ 2



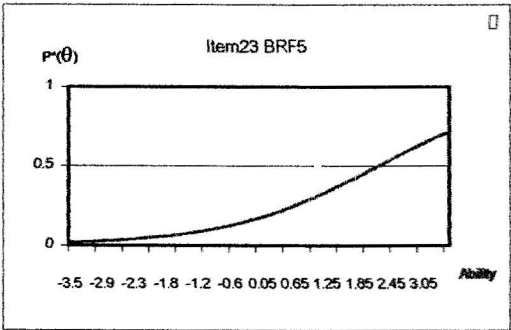
ขอบเขตลำดับขั้นที่ 3



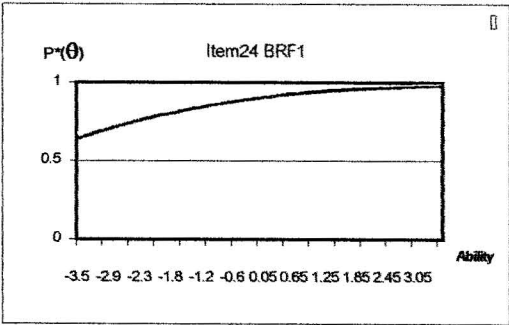
ขอบเขตลำดับขั้นที่ 4



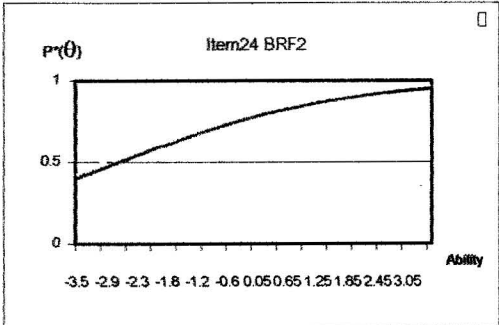
ขอบเขตลำดับขั้นที่ 5



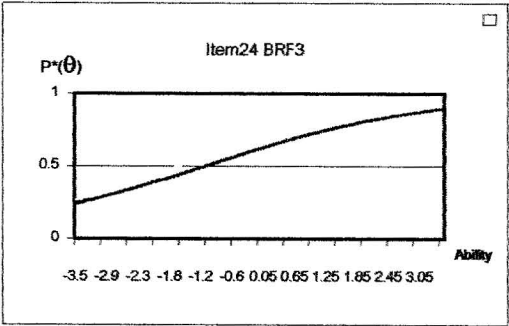
ข้อ 24 ขอบเขตลำดับขั้นที่ 1



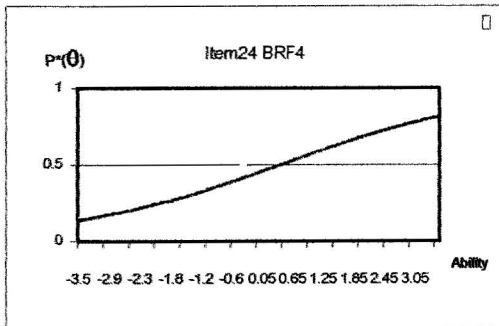
ขอบเขตลำดับขั้นที่ 2



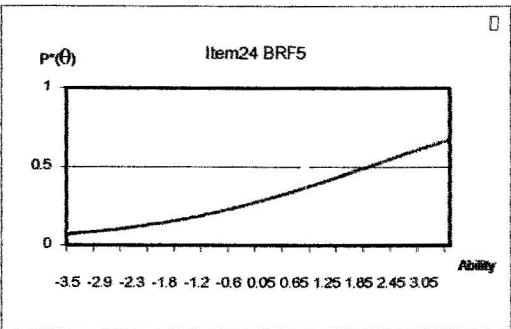
ขอบเขตลำดับขั้นที่ 3



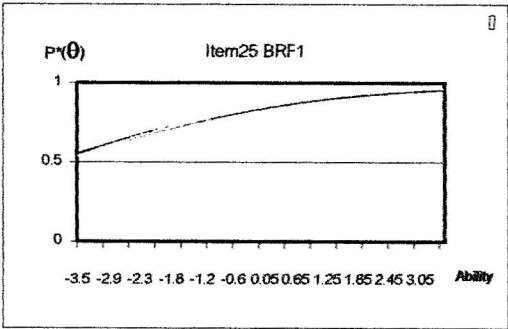
ขอบเขตลำดับขั้นที่ 4



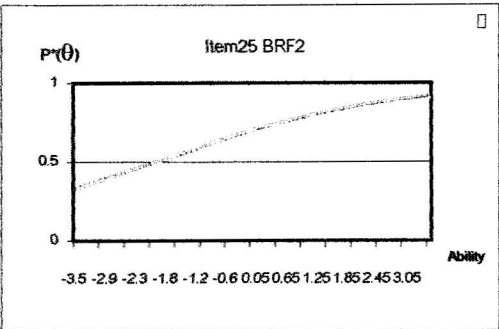
ขอบเขตลำดับขั้นที่ 5



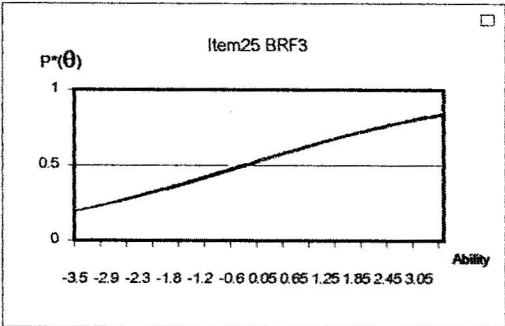
ข้อ 25 ขอบเขตลำดับขั้นที่ 1



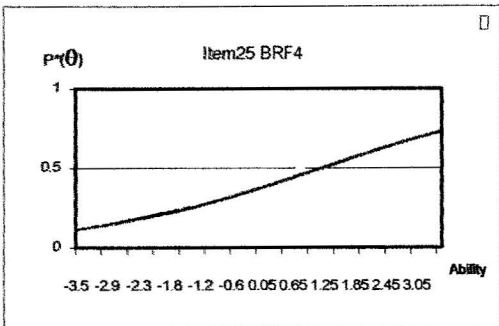
ข้อ 25 ขอบเขตลำดับขั้นที่ 2



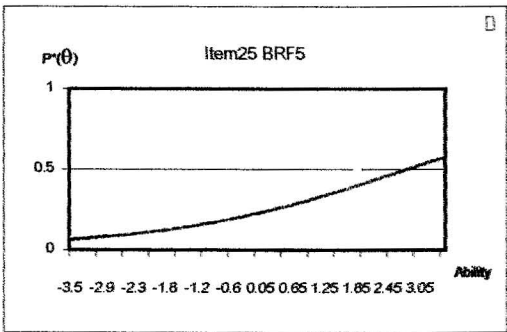
ข้อ 25 ขอบเขตลำดับขั้นที่ 3



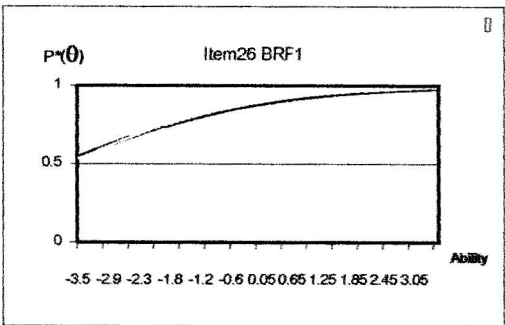
ข้อ 25 ขอบเขตลำดับขั้นที่ 4



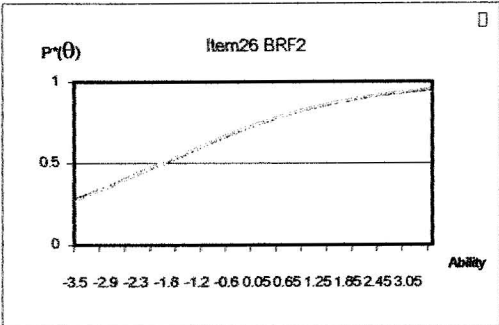
ข้อ 25 ขอบเขตลำดับขั้นที่ 5



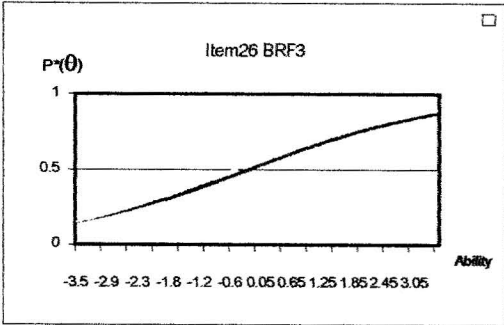
ข้อ 26 ขอบเขตลำดับขั้นที่ 1



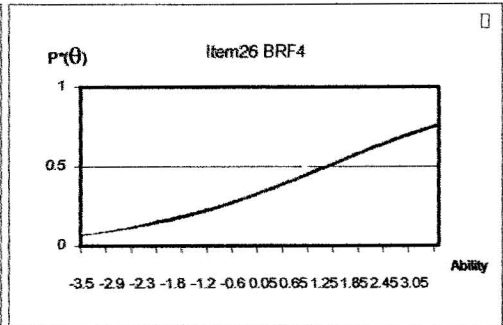
ข้อ 26 ขอบเขตลำดับขั้นที่ 2



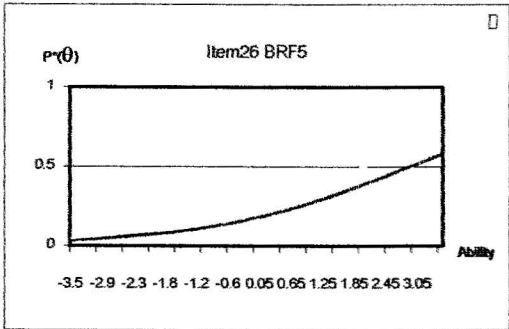
ขอบเขตลำดับขั้นที่ 3



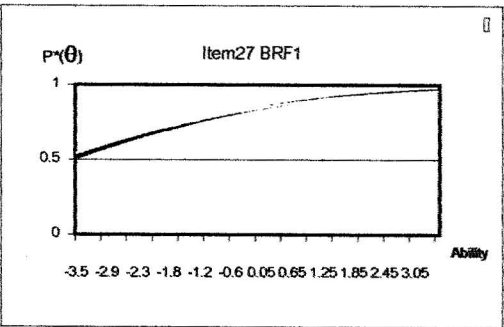
ขอบเขตลำดับขั้นที่ 4



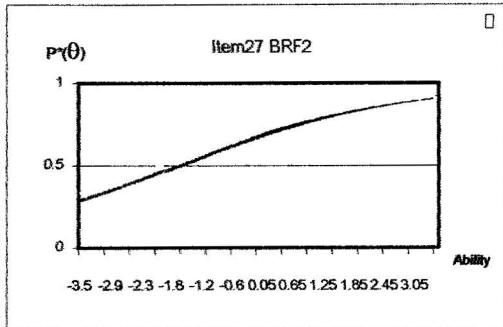
ขอบเขตลำดับขั้นที่ 5



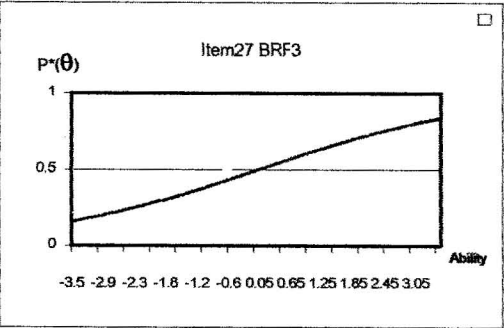
ข้อ 27 ขอบเขตลำดับขั้นที่ 1



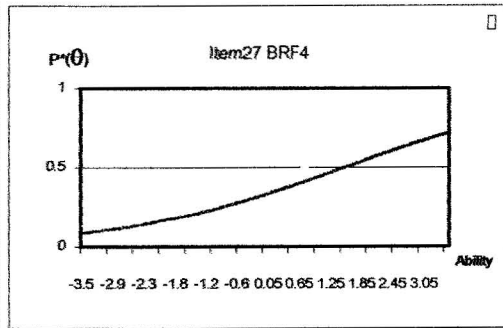
ขอบเขตลำดับขั้นที่ 2



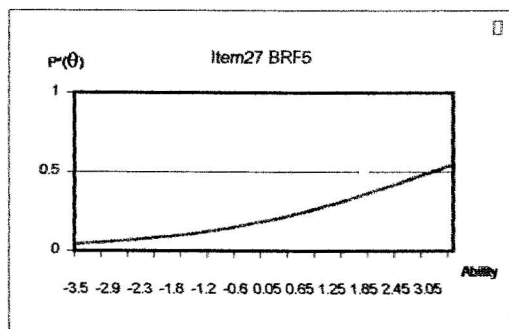
ขอบเขตลำดับขั้นที่ 3



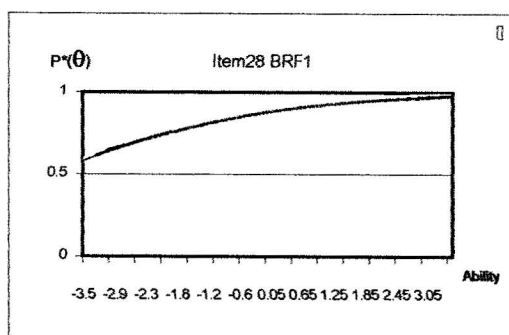
ขอบเขตลำดับขั้นที่ 4



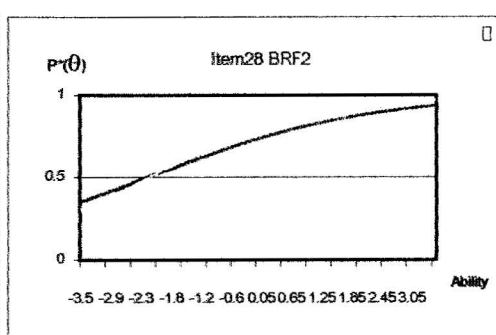
ขอบเขตลำดับขั้นที่ 5



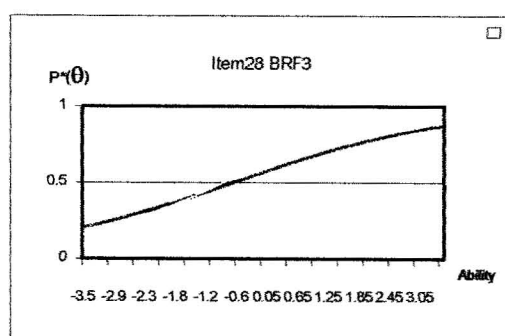
ข้อ 28 ขอบเขตลำดับขั้นที่ 1



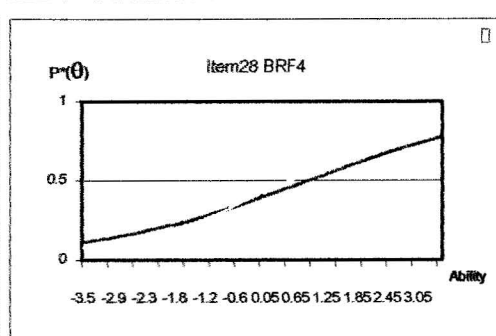
ขอบเขตลำดับขั้นที่ 2



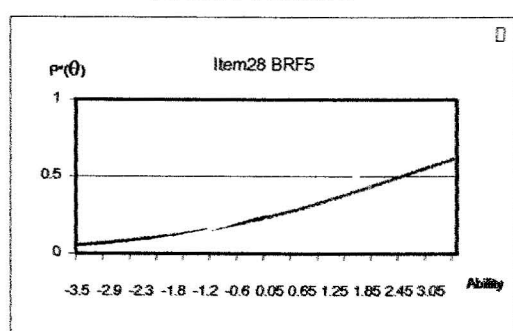
ขอบเขตลำดับขั้นที่ 3



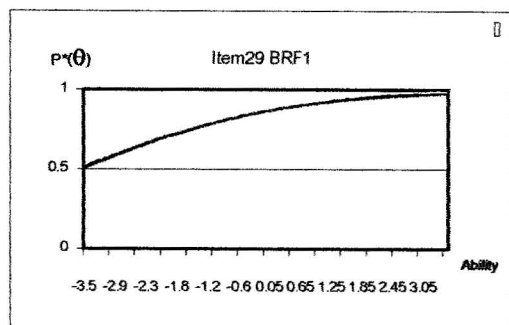
ขอบเขตลำดับขั้นที่ 4



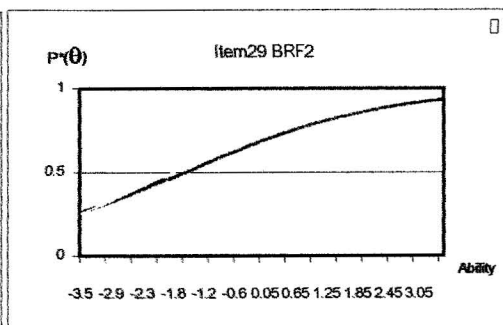
ขอบเขตลำดับขั้นที่ 5



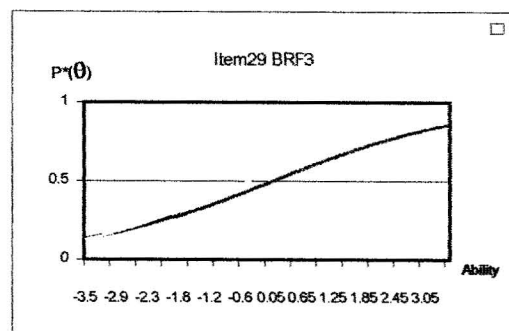
ข้อ 29 ขอบเขตลำดับขั้นที่ 1



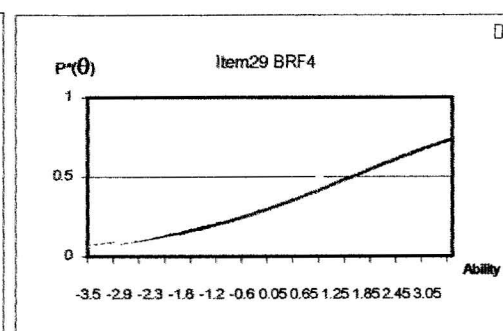
ขอบเขตลำดับขั้นที่ 2



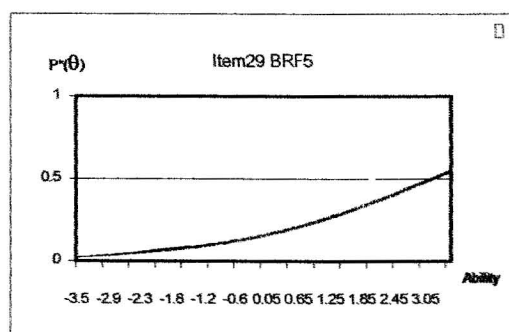
ขอบเขตลำดับขั้นที่ 3



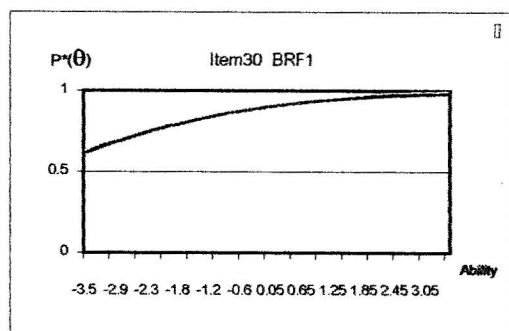
ขอบเขตลำดับขั้นที่ 4



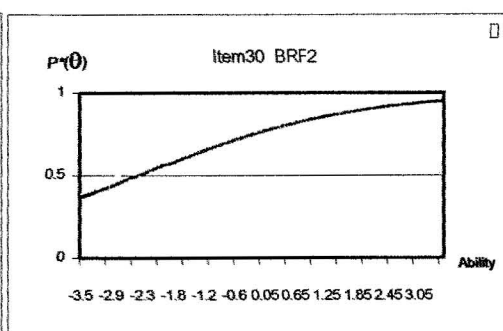
ขอบเขตลำดับขั้นที่ 5



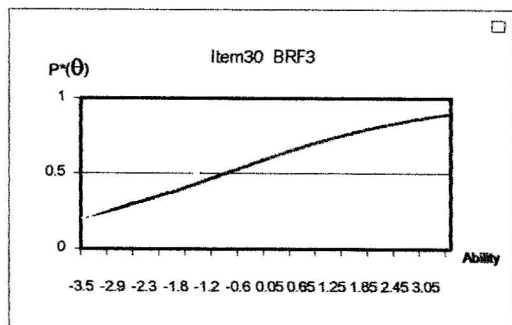
ข้อ 30 ขอบเขตลำดับขั้นที่ 1



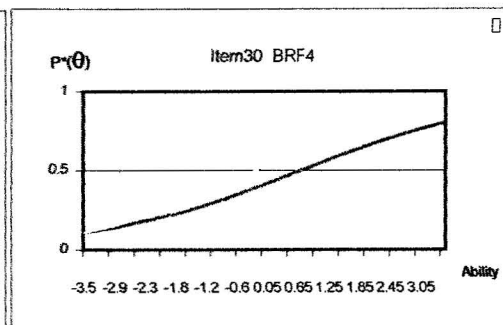
ขอบเขตลำดับขั้นที่ 2



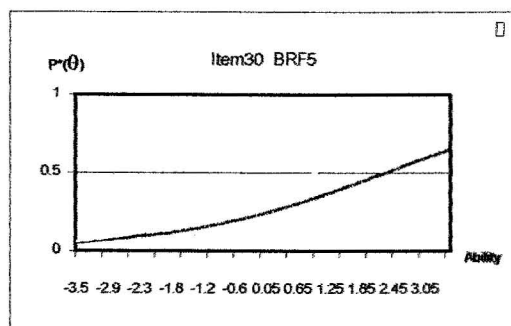
ขอบเขตลำดับขั้นที่ 3



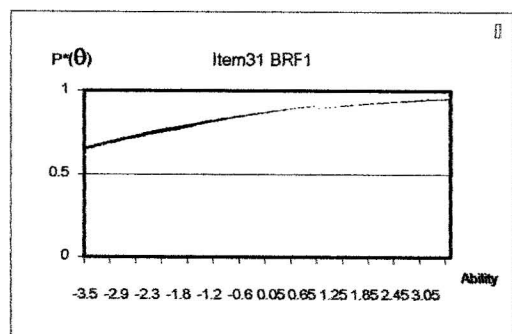
ขอบเขตลำดับขั้นที่ 4



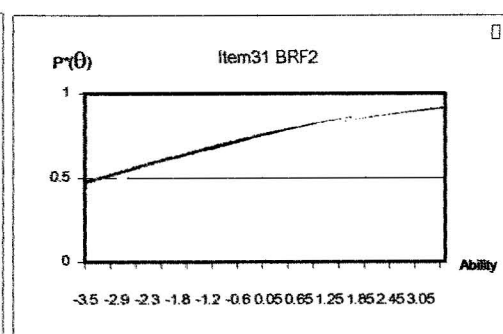
ขอบเขตลำดับขั้นที่ 5



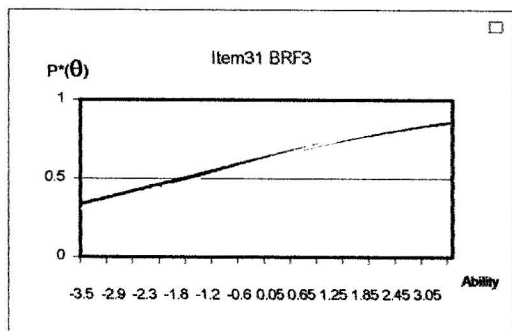
ข้อ 31 ขอบเขตลำดับขั้นที่ 1



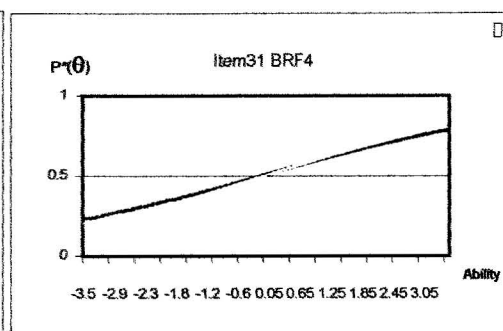
ขอบเขตลำดับขั้นที่ 2



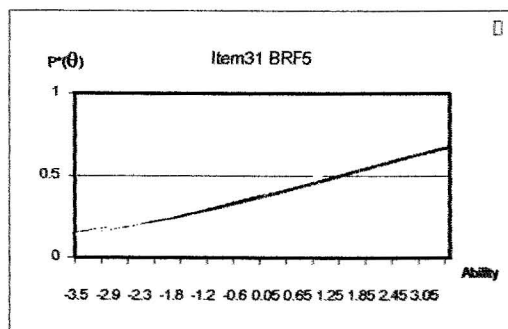
ขอบเขตลำดับขั้นที่ 3



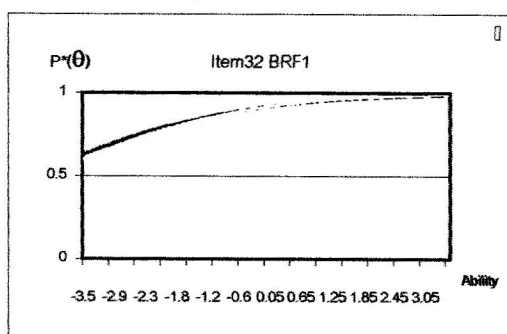
ขอบเขตลำดับขั้นที่ 4



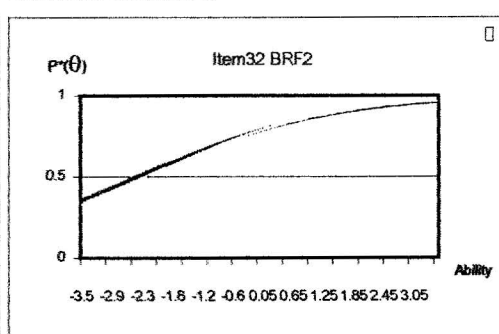
ขอบเขตลำดับขั้นที่ 5



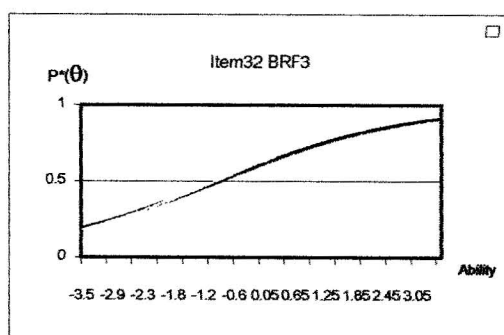
ข้อ 32 ขอบเขตลำดับขั้นที่ 1



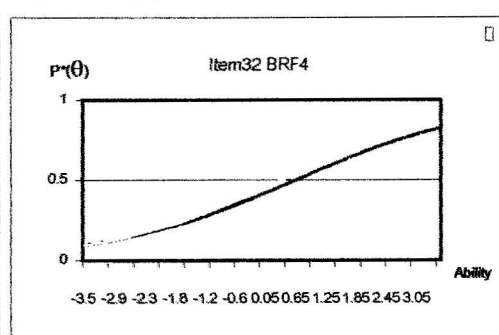
ขอบเขตลำดับขั้นที่ 2



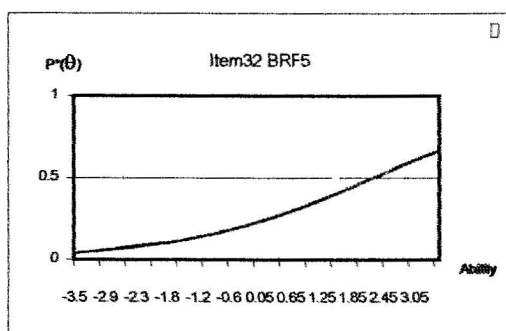
ขอบเขตลำดับขั้นที่ 3



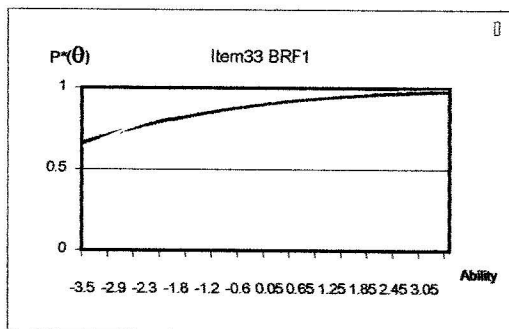
ขอบเขตลำดับขั้นที่ 4



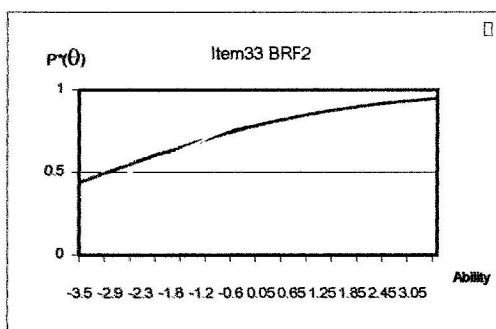
ขอบเขตลำดับขั้นที่ 5



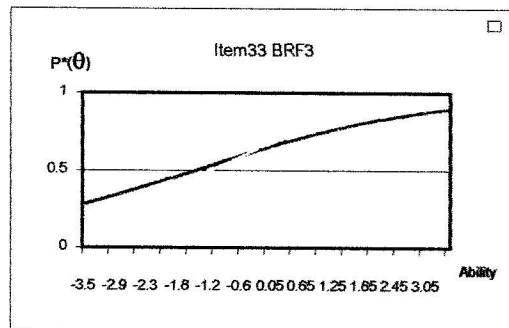
ข้อ 33 ขอบเขตลำดับขั้นที่ 1



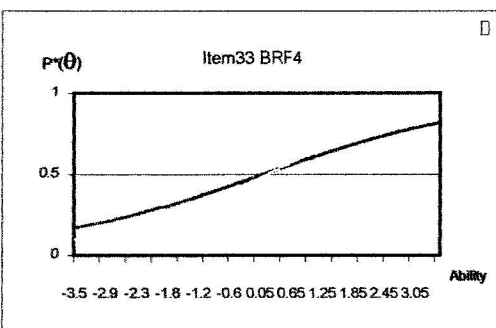
ขอบเขตลำดับขั้นที่ 2



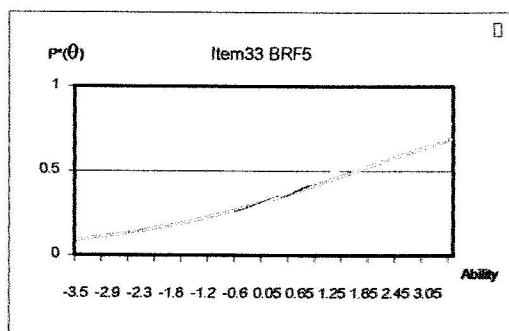
ขอบเขตลำดับขั้นที่ 3



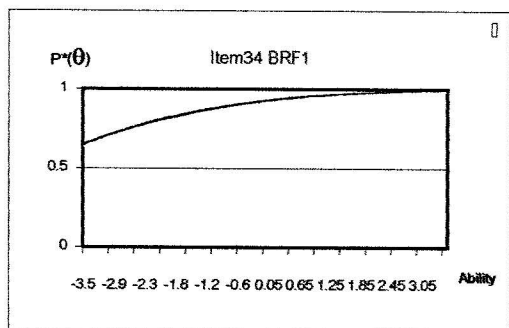
ขอบเขตลำดับขั้นที่ 4



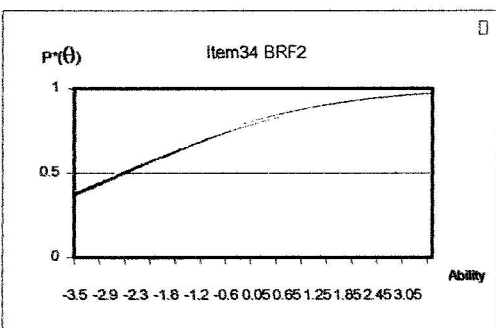
ขอบเขตลำดับขั้นที่ 5



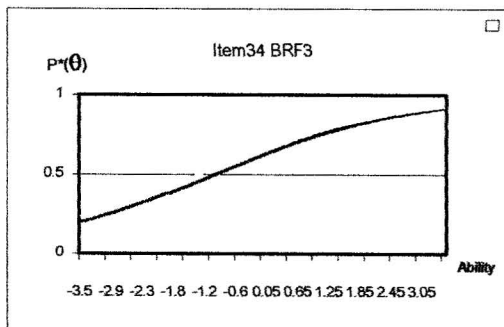
ข้อ 34 ขอบเขตลำดับขั้นที่ 1



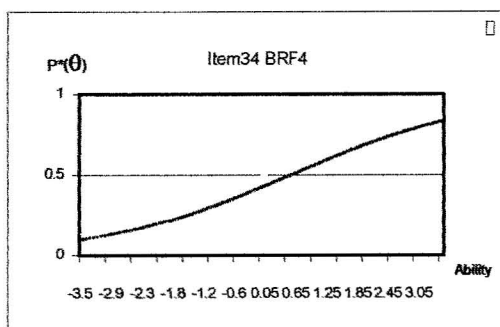
ขอบเขตลำดับขั้นที่ 2



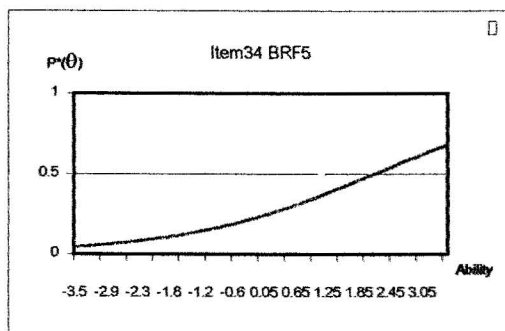
ขอบเขตลำดับขั้นที่ 3



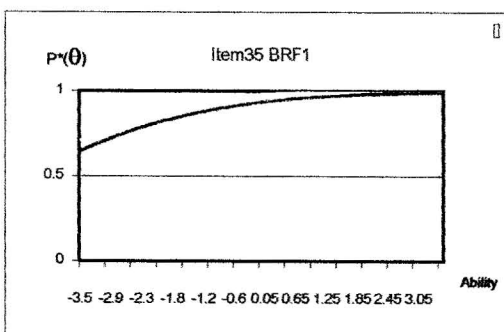
ขอบเขตลำดับขั้นที่ 4



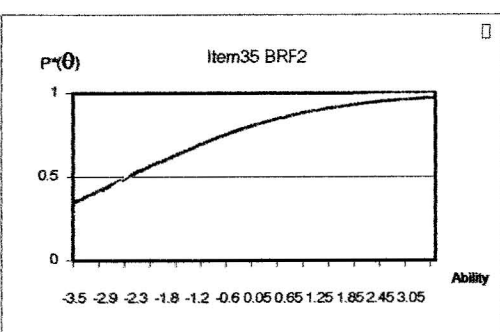
ขอบเขตลำดับขั้นที่ 5



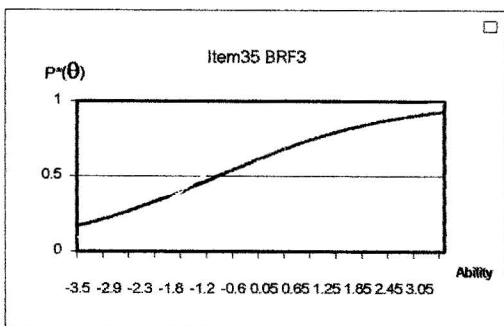
ข้อ 35 ขอบเขตลำดับขั้นที่ 1



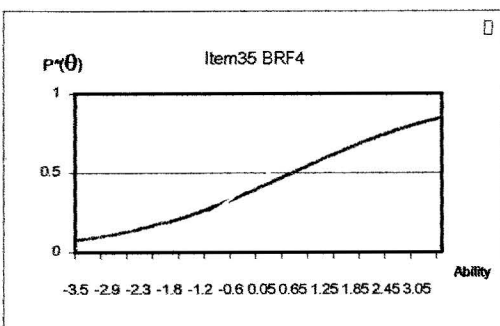
ขอบเขตลำดับขั้นที่ 2



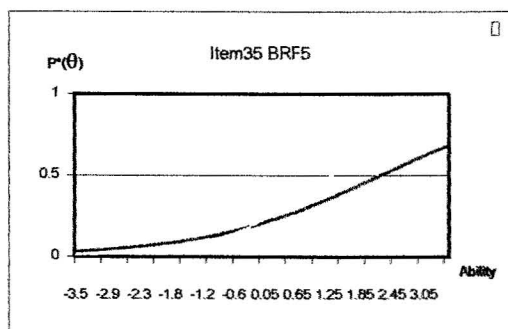
ขอบเขตลำดับขั้นที่ 3



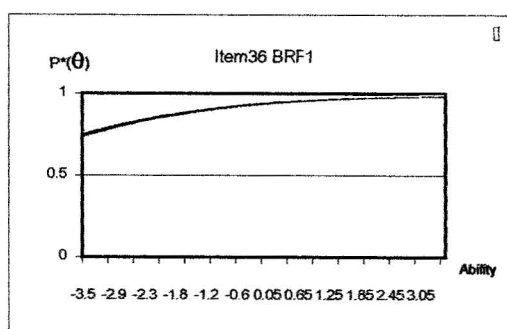
ขอบเขตลำดับขั้นที่ 4



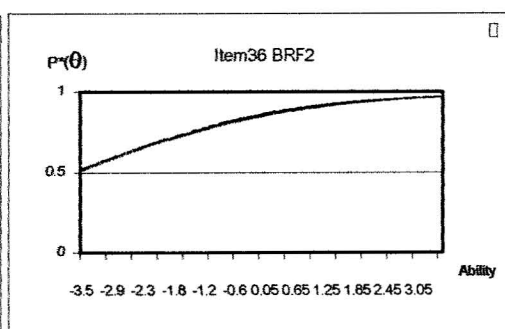
ขอบเขตลำดับขั้นที่ 5



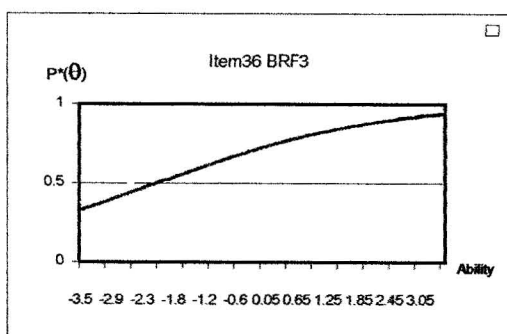
ข้อ 36 ขอบเขตลำดับขั้นที่ 1



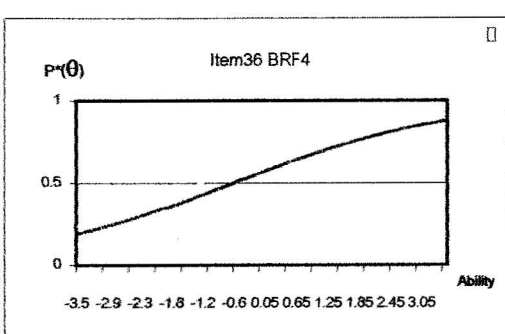
ขอบเขตลำดับขั้นที่ 2



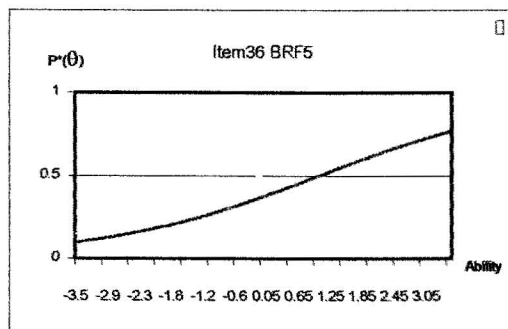
ขอบเขตลำดับขั้นที่ 3



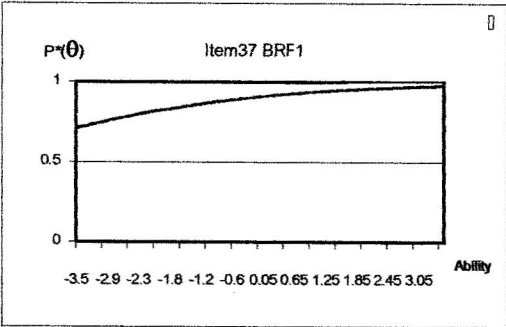
ขอบเขตลำดับขั้นที่ 4



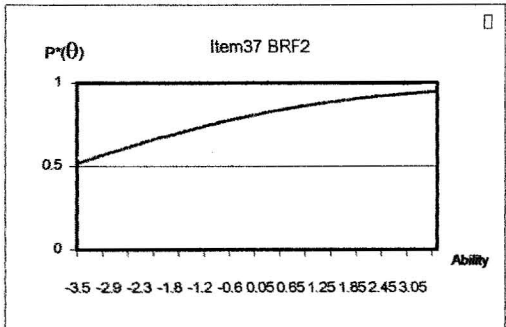
ขอบเขตลำดับขั้นที่ 5



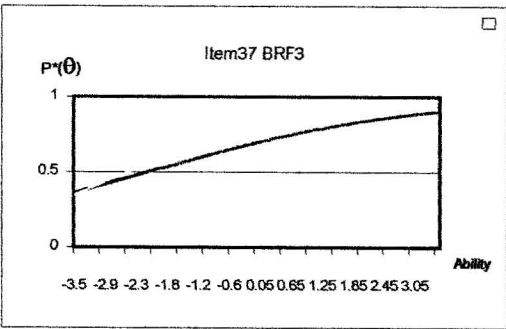
ข้อ 37 ขอบเขตลำดับขั้นที่ 1



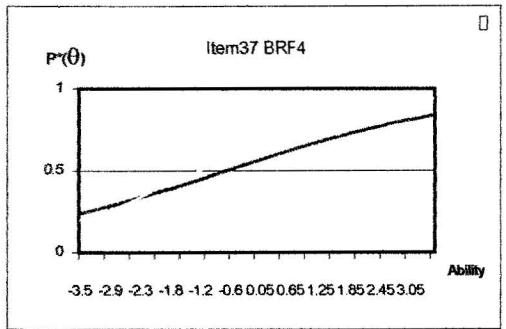
ข้อ 37 ขอบเขตลำดับขั้นที่ 2



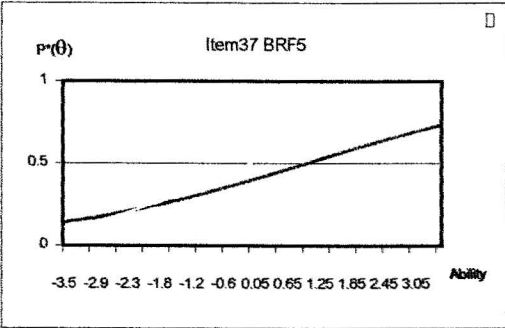
ข้อ 37 ขอบเขตลำดับขั้นที่ 3



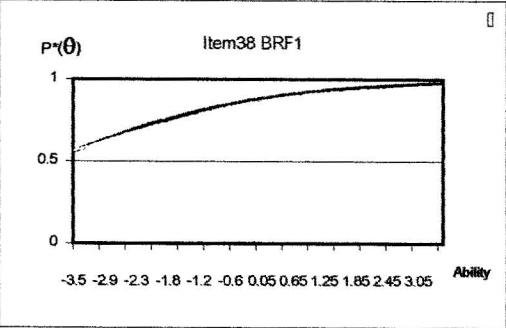
ข้อ 37 ขอบเขตลำดับขั้นที่ 4



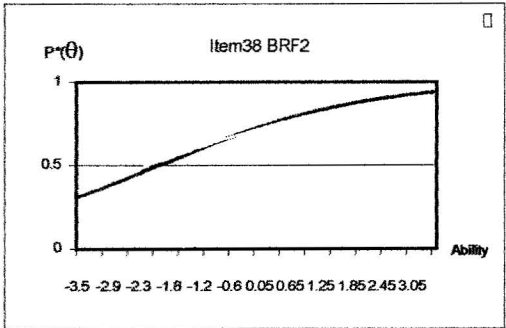
ข้อ 37 ขอบเขตลำดับขั้นที่ 5



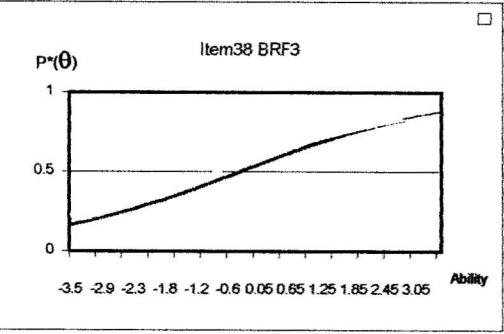
ข้อ 38 ขอบเขตลำดับขั้นที่ 1



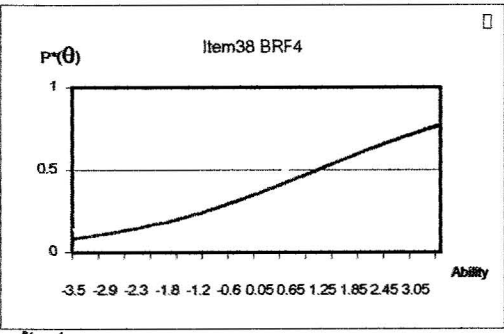
ข้อ 38 ขอบเขตลำดับขั้นที่ 2



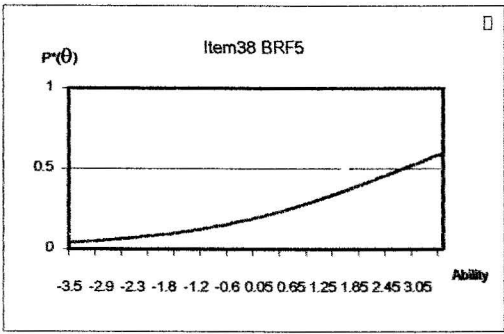
ขอบเขตลำดับขั้นที่ 3



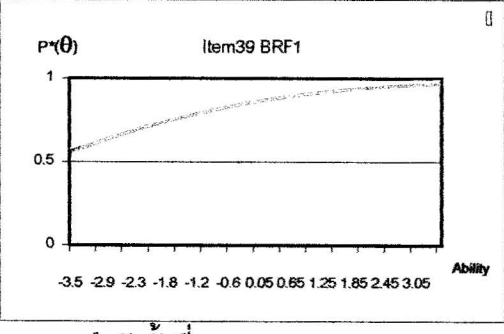
ขอบเขตลำดับขั้นที่ 4



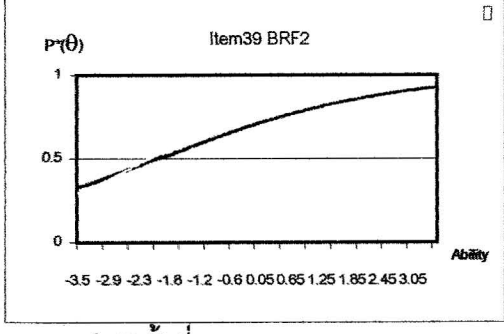
ขอบเขตลำดับขั้นที่ 5



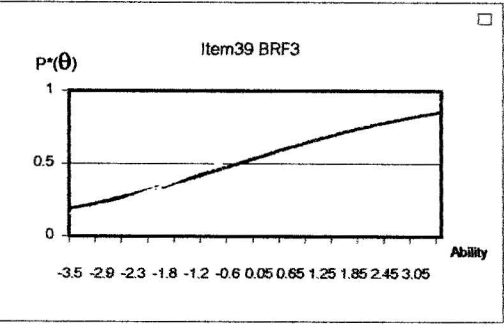
ข้อ 39 ขอบเขตลำดับขั้นที่ 1



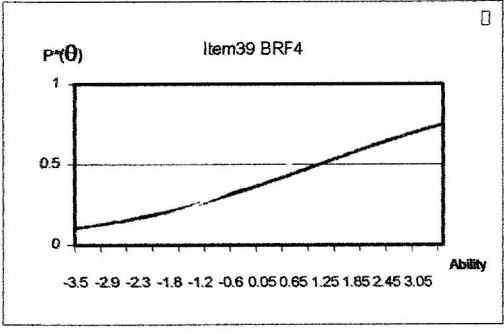
ขอบเขตลำดับขั้นที่ 2



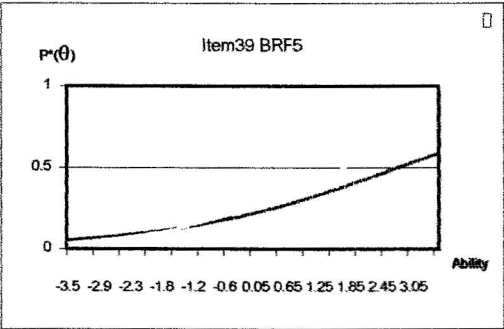
ขอบเขตลำดับขั้นที่ 3



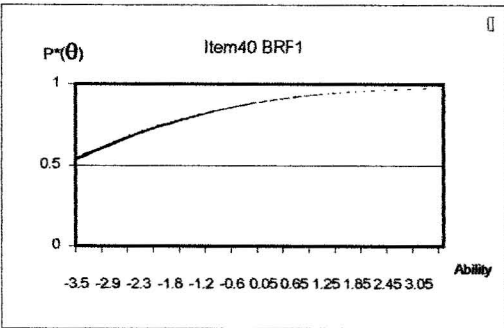
ขอบเขตลำดับขั้นที่ 4



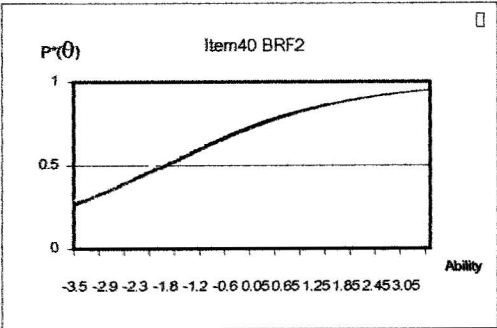
ขอบเขตลำดับขั้นที่ 5



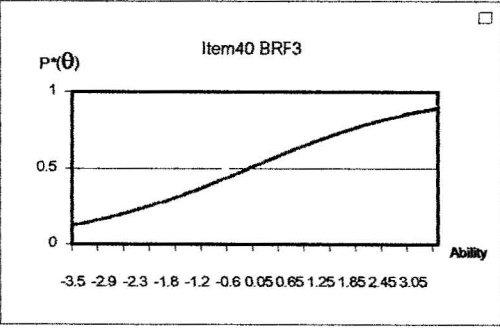
ข้อ 40 ขอบเขตลำดับขั้นที่ 1



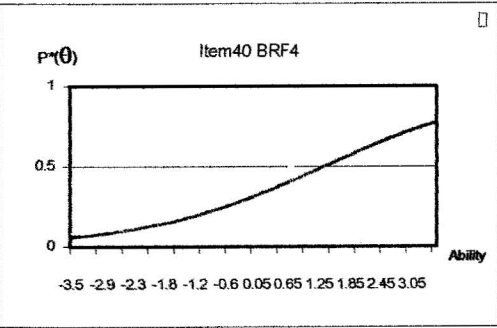
ขอบเขตลำดับขั้นที่ 2



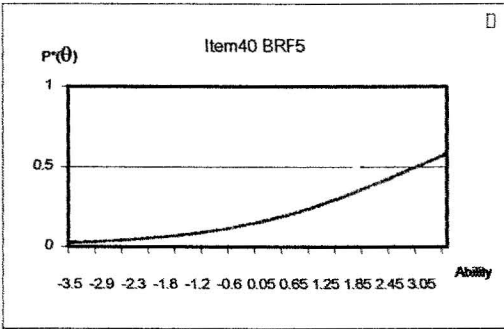
ขอบเขตลำดับขั้นที่ 3



ขอบเขตลำดับขั้นที่ 4



ขอบเขตลำดับขั้นที่ 5

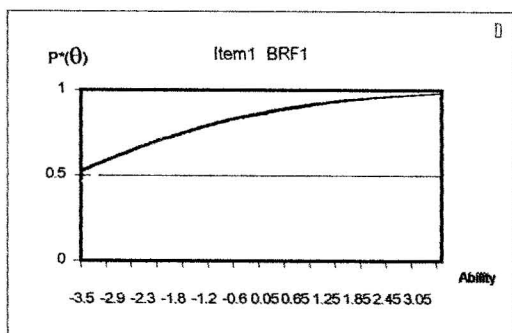


ฟังก์ชันผลการตอบข้อสอบในขอบเขตลำดับชั้น (BRF) ที่วิเคราะห์ด้วยวิธี GPCM ในแบบทดสอบขนาด 30 ข้อ
เมื่อกลุ่มตัวอย่างเพศชายต่อเพศหญิง เท่ากับ 1,000 คน ต่อ 1,000 คน

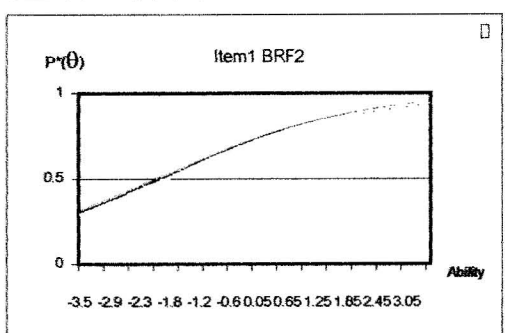
ชาย

หญิง

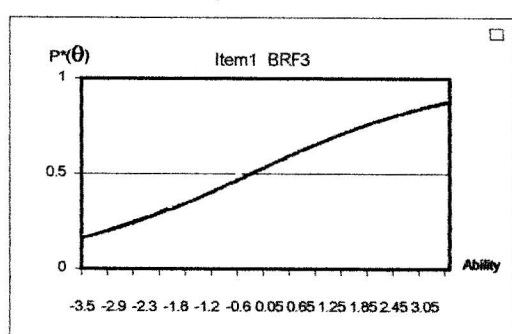
ข้อ 1 ขอบเขตลำดับชั้นที่ 1



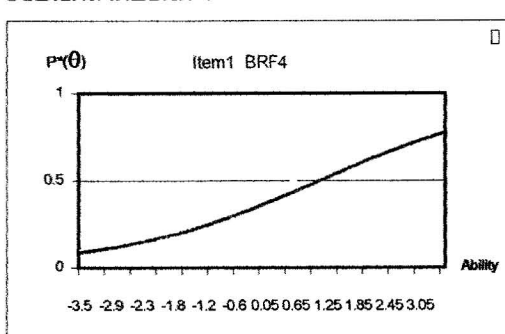
ขอบเขตลำดับชั้นที่ 2



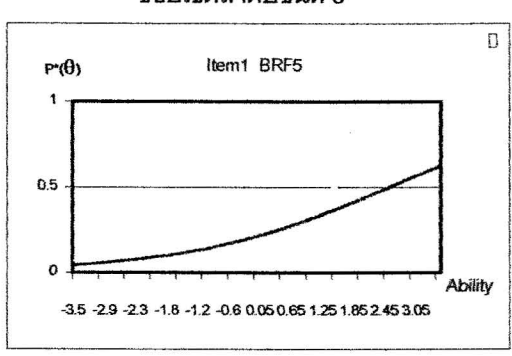
ขอบเขตลำดับชั้นที่ 3



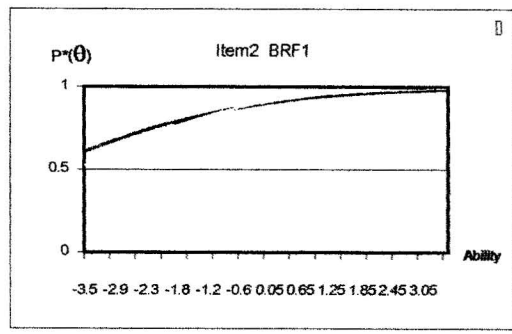
ขอบเขตลำดับชั้นที่ 4



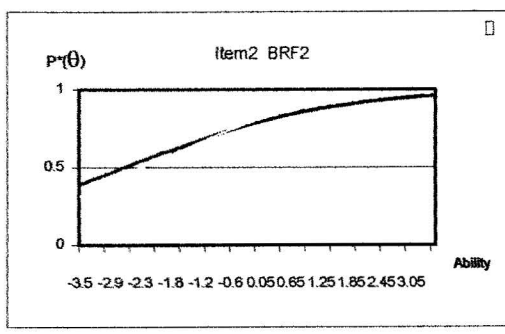
ขอบเขตลำดับชั้นที่ 5



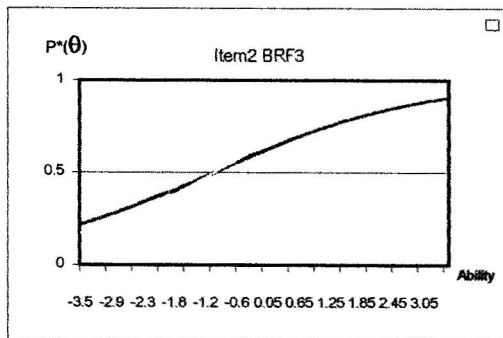
ข้อ 2 ขอบเขตลำดับชั้นที่ 1



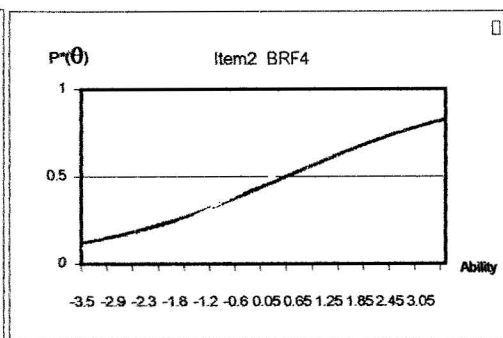
ขอบเขตลำดับชั้นที่ 2



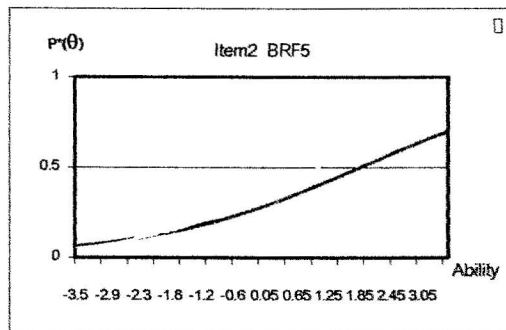
ขอบเขตลำดับขั้นที่ 3



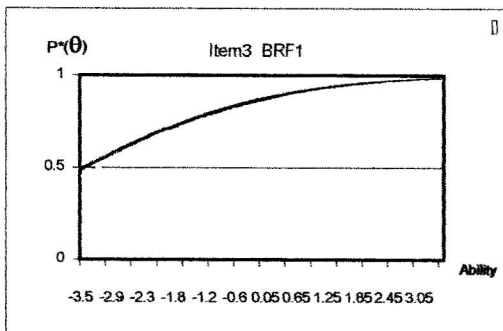
ขอบเขตลำดับขั้นที่ 4



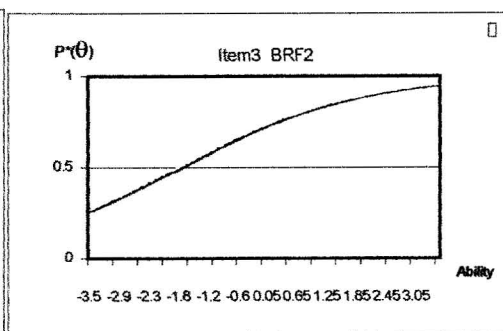
ขอบเขตลำดับขั้นที่ 5



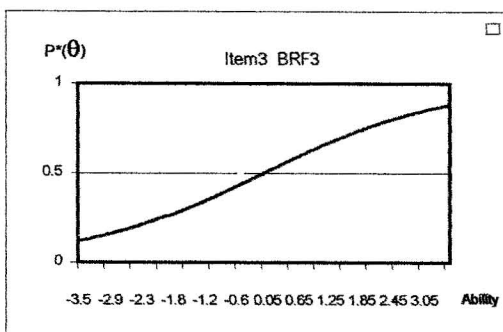
ข้อ 3 ขอบเขตลำดับขั้นที่ 1



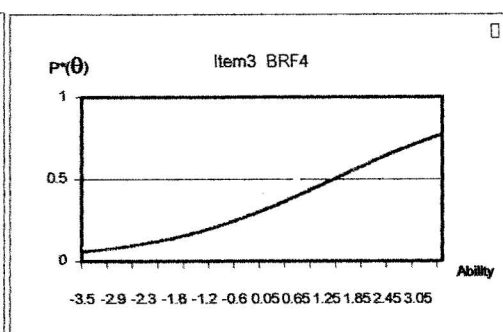
ขอบเขตลำดับขั้นที่ 2



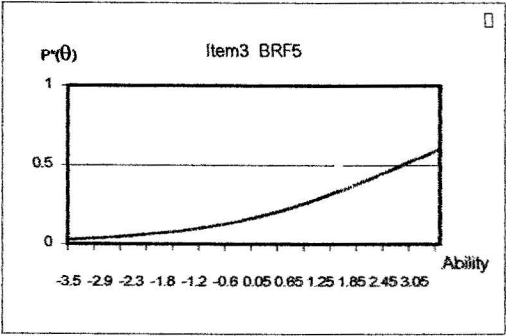
ขอบเขตลำดับขั้นที่ 3



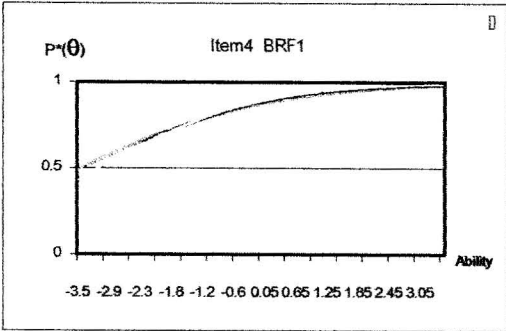
ขอบเขตลำดับขั้นที่ 4



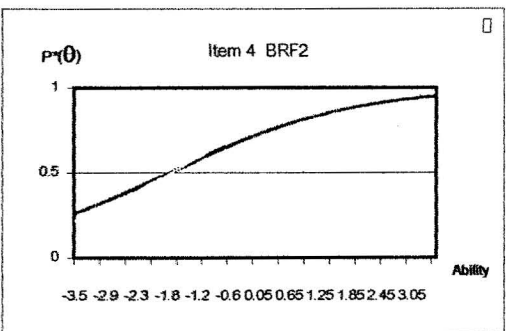
ขอบเขตลำดับขั้นที่ 5



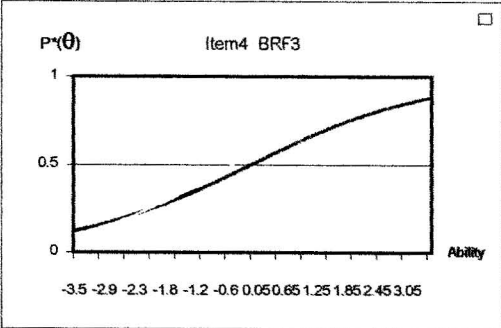
ข้อ 4 ขอบเขตลำดับขั้นที่ 1



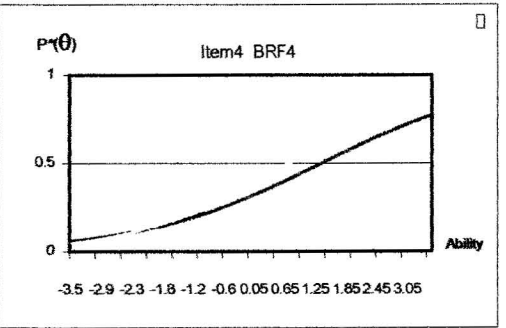
ขอบเขตลำดับขั้นที่ 2



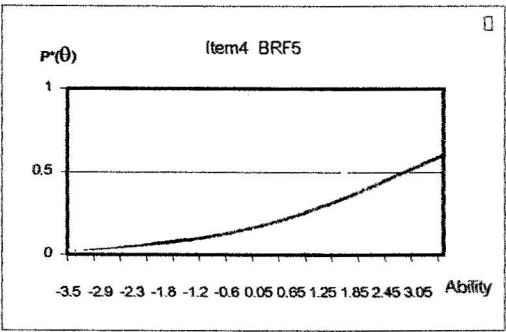
ขอบเขตลำดับขั้นที่ 3



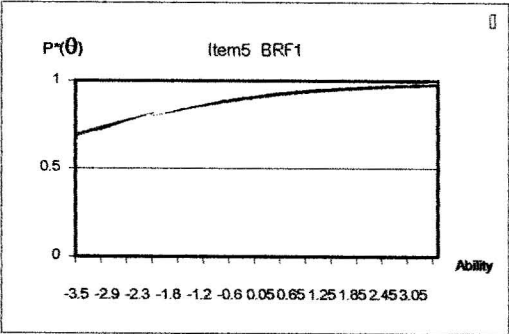
ขอบเขตลำดับขั้นที่ 4



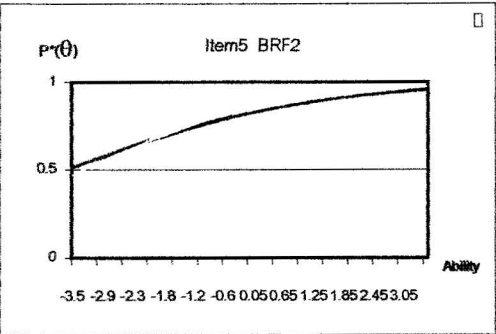
ขอบเขตลำดับขั้นที่ 5



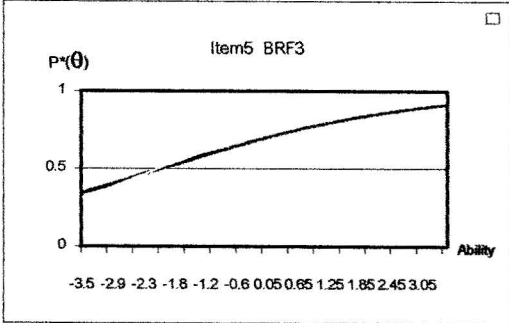
ข้อ 5 ขอบเขตลำดับขั้นที่ 1



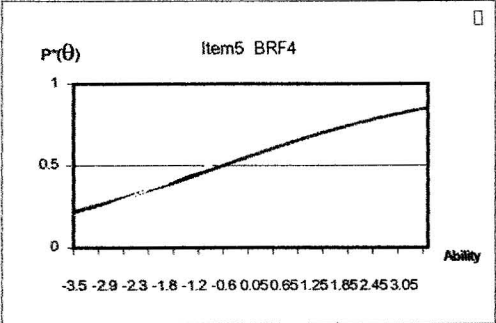
ข้อ 5 ขอบเขตลำดับขั้นที่ 2



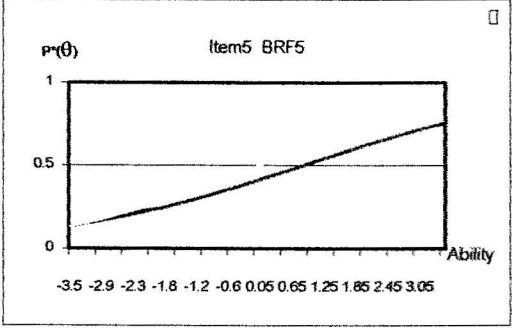
ข้อ 5 ขอบเขตลำดับขั้นที่ 3



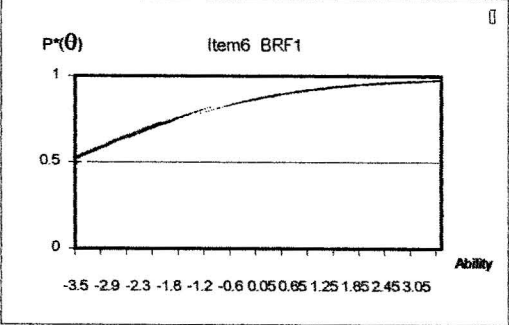
ข้อ 5 ขอบเขตลำดับขั้นที่ 4



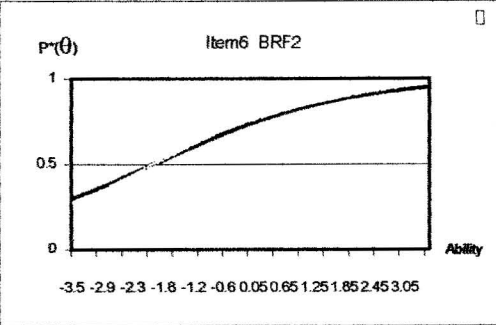
ข้อ 5 ขอบเขตลำดับขั้นที่ 5



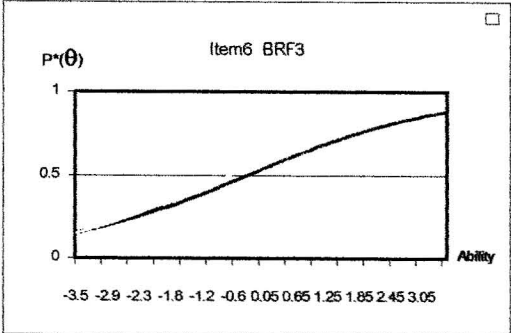
ข้อ 6 ขอบเขตลำดับขั้นที่ 1



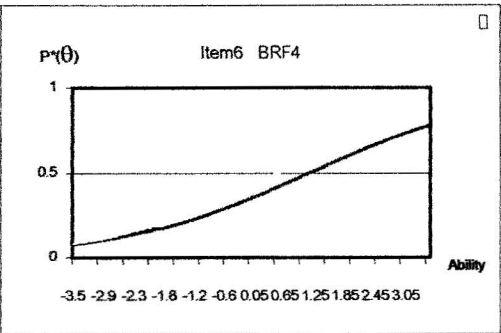
ข้อ 6 ขอบเขตลำดับขั้นที่ 2



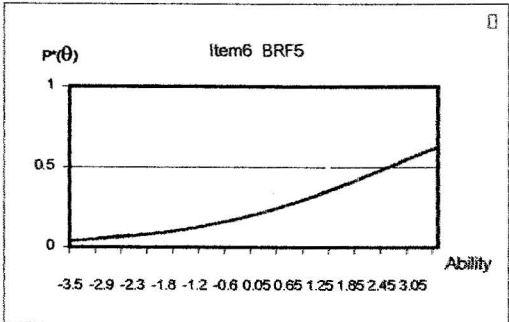
ข้อ 10 ขอบเขตลำดับขั้นที่ 3



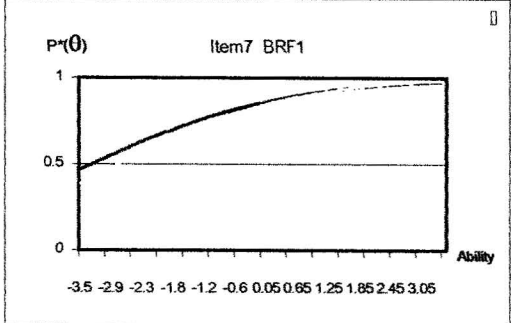
ข้อ 10 ขอบเขตลำดับขั้นที่ 4



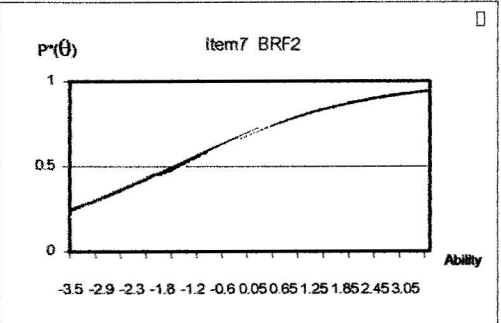
ข้อ 10 ขอบเขตลำดับขั้นที่ 5



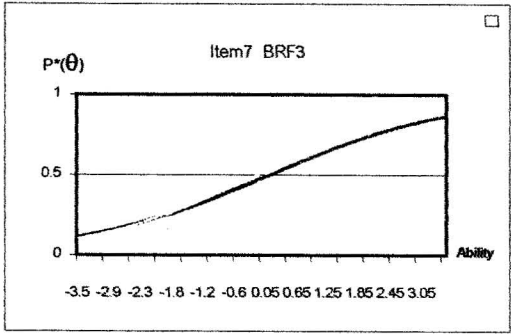
ข้อ 10 ขอบเขตลำดับขั้นที่ 1



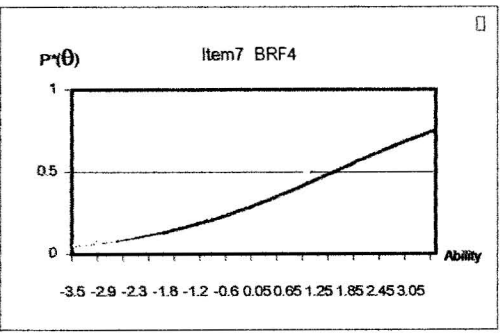
ข้อ 10 ขอบเขตลำดับขั้นที่ 2



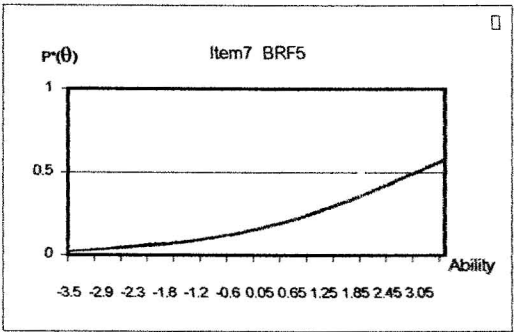
ข้อ 10 ขอบเขตลำดับขั้นที่ 3



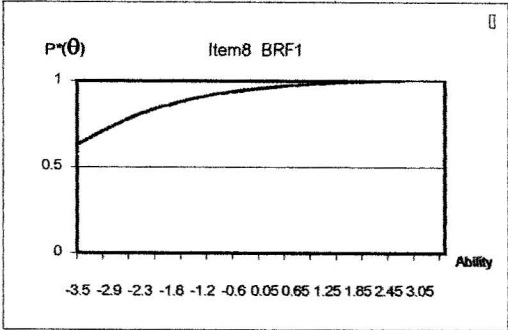
ข้อ 10 ขอบเขตลำดับขั้นที่ 4



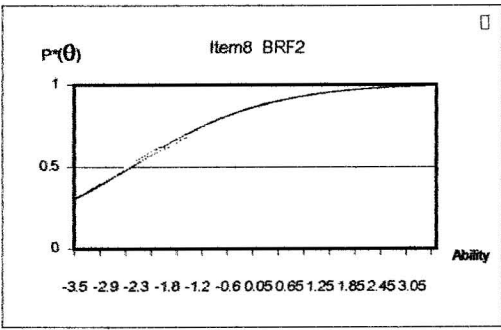
ขอบเขตลำดับขั้นที่ 5



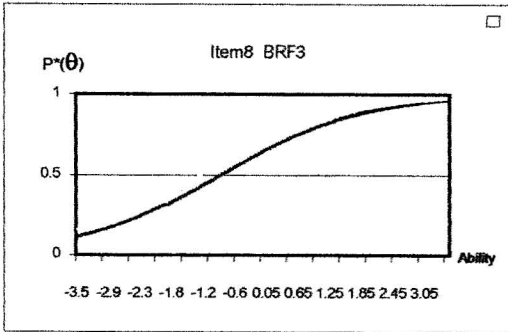
ข้อ 8 ขอบเขตลำดับขั้นที่ 1



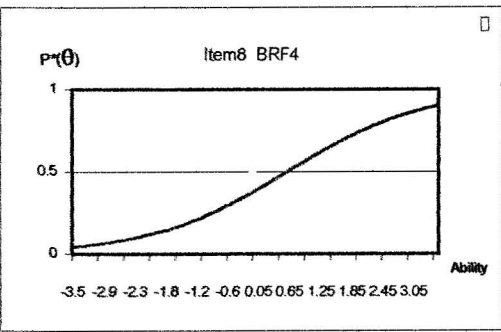
ขอบเขตลำดับขั้นที่ 2



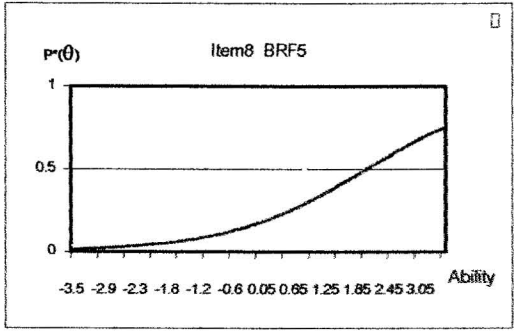
ขอบเขตลำดับขั้นที่ 3



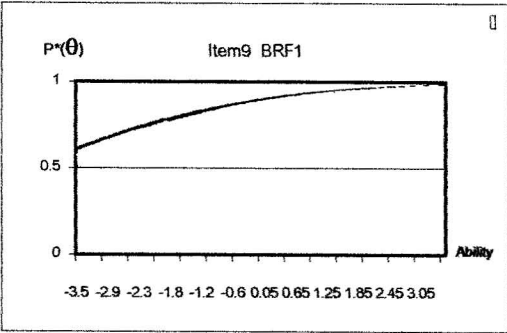
ขอบเขตลำดับขั้นที่ 4



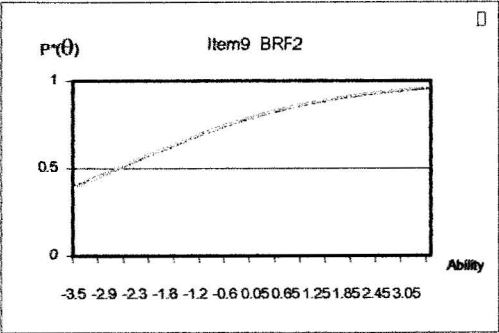
ขอบเขตลำดับขั้นที่ 5



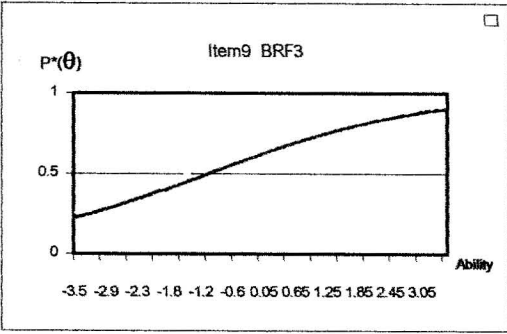
ข้อ 9 ขอบเขตลำดับขั้นที่ 1



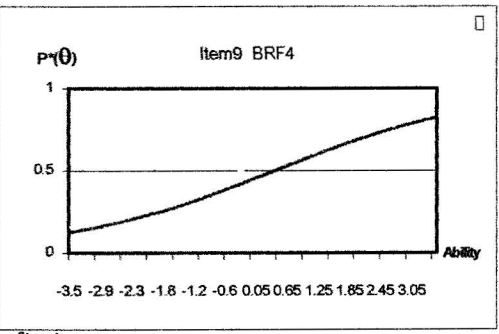
ขอบเขตลำดับขั้นที่ 2



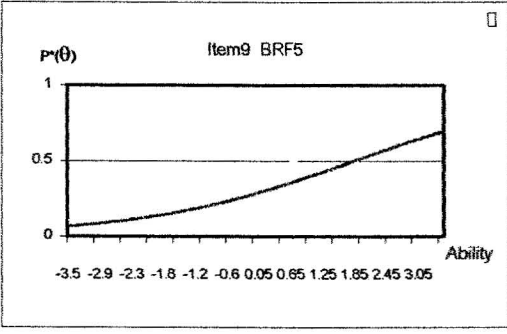
ขอบเขตลำดับขั้นที่ 3



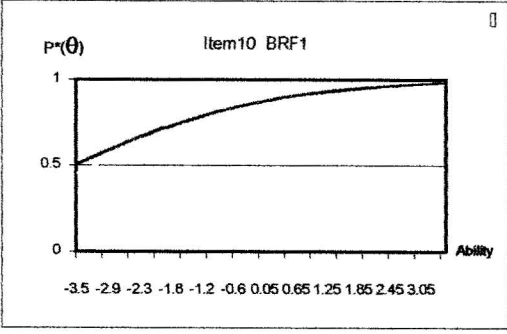
ขอบเขตลำดับขั้นที่ 4



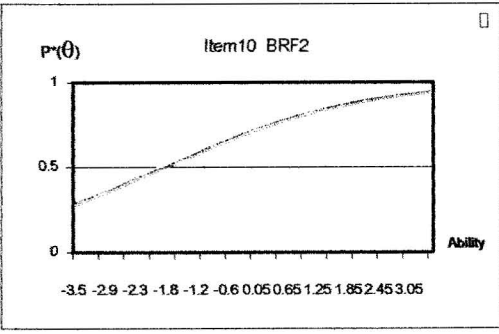
ขอบเขตลำดับขั้นที่ 5



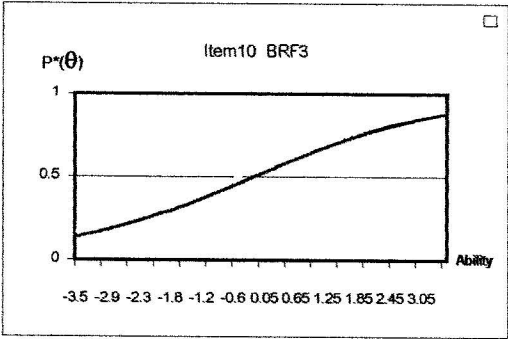
ข้อ 10 ขอบเขตลำดับขั้นที่ 1



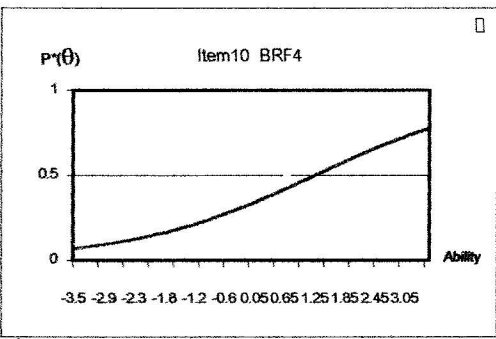
ขอบเขตลำดับขั้นที่ 2



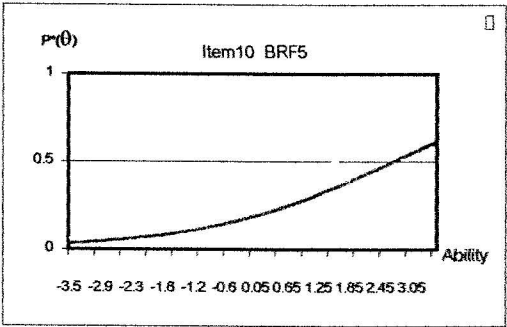
ขอบเขตลำดับขั้นที่ 3



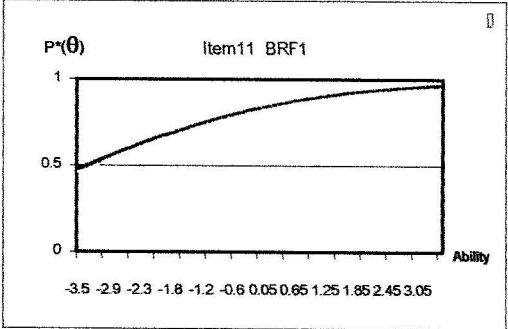
ขอบเขตลำดับขั้นที่ 4



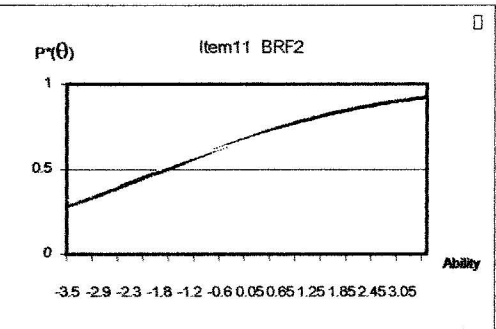
ขอบเขตลำดับขั้นที่ 5



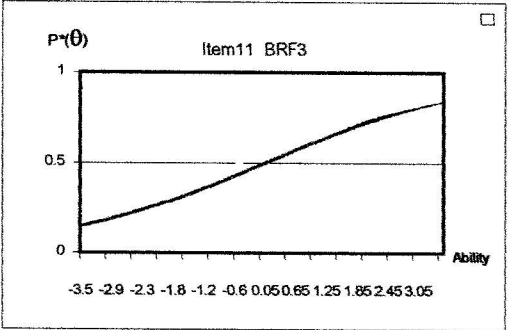
ข้อ 11 ขอบเขตลำดับขั้นที่ 1



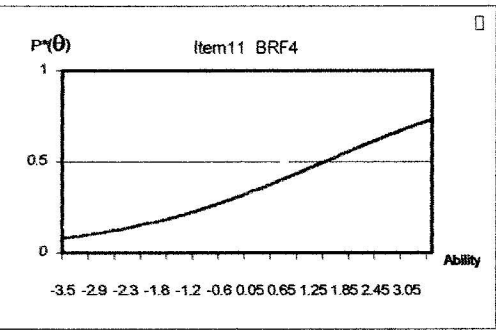
ขอบเขตลำดับขั้นที่ 2



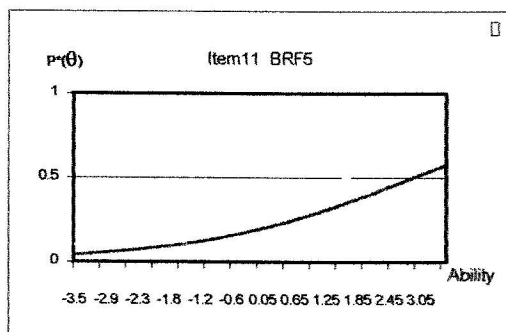
ขอบเขตลำดับขั้นที่ 3



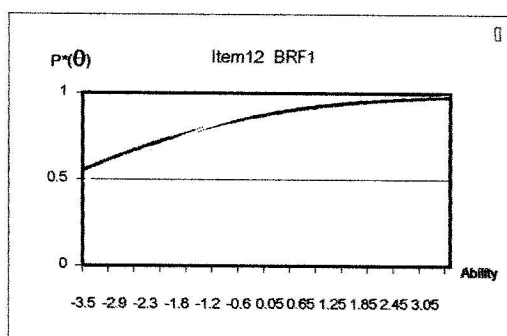
ขอบเขตลำดับขั้นที่ 4



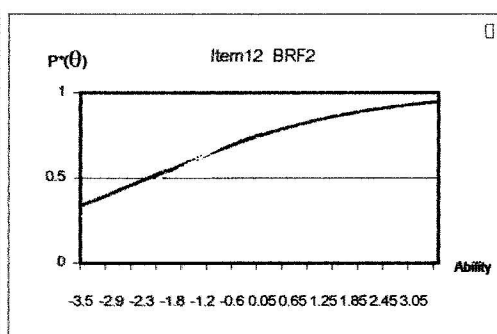
ขอบเขตลำดับขั้นที่ 5



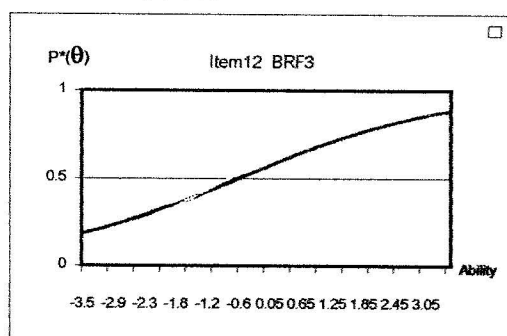
ข้อ 12 ขอบเขตลำดับขั้นที่ 1



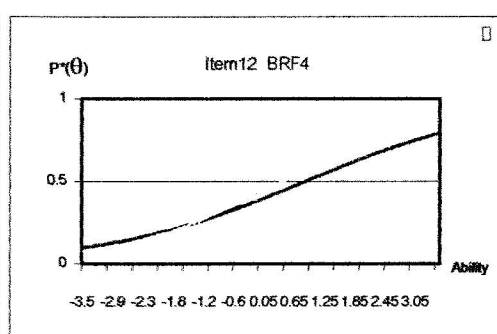
ขอบเขตลำดับขั้นที่ 2



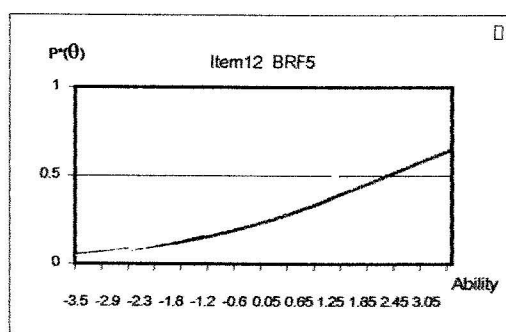
ขอบเขตลำดับขั้นที่ 3



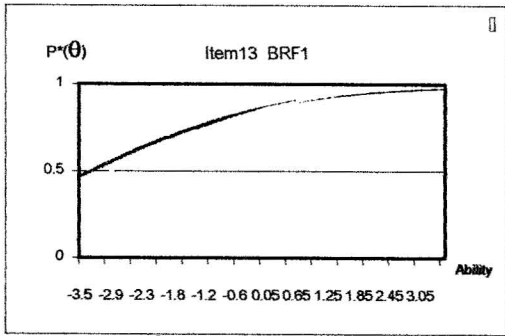
ขอบเขตลำดับขั้นที่ 4



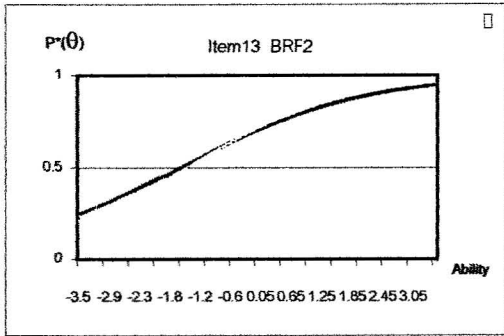
ขอบเขตลำดับขั้นที่ 5



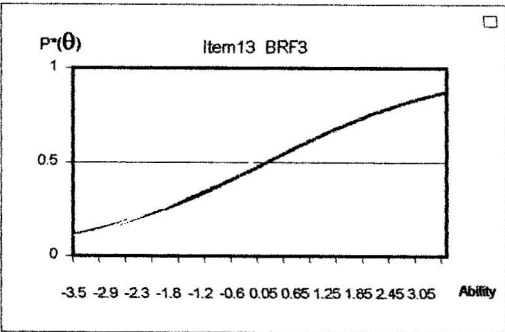
ข้อ 13 ขอบเขตลำดับขั้นที่ 1



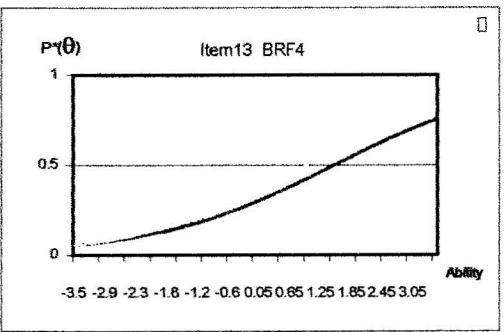
ขอบเขตลำดับขั้นที่ 2



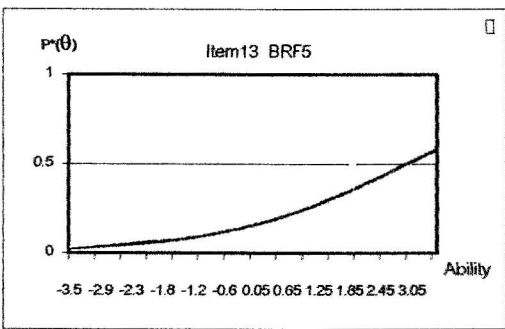
ขอบเขตลำดับขั้นที่ 3



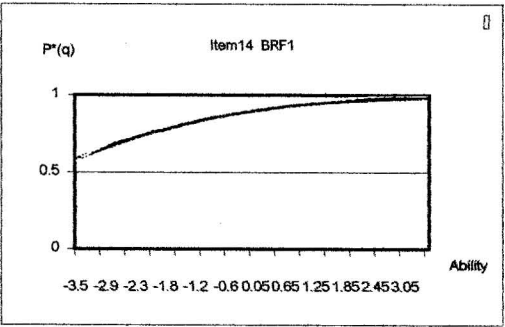
ขอบเขตลำดับขั้นที่ 4



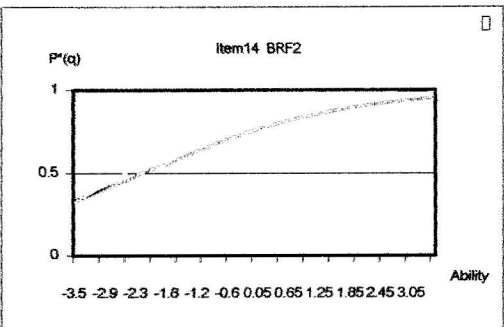
ขอบเขตลำดับขั้นที่ 5



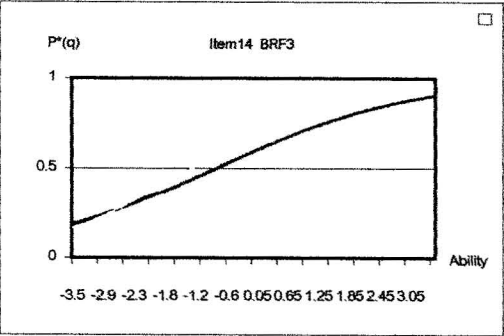
ข้อ 14 ขอบเขตลำดับขั้นที่ 1



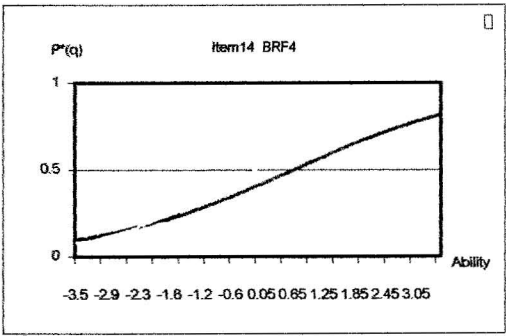
ขอบเขตลำดับขั้นที่ 2



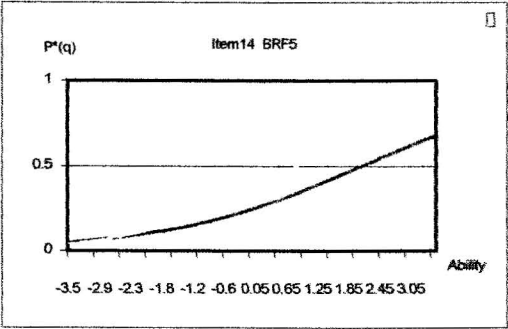
ขอบเขตลำดับขั้นที่ 3



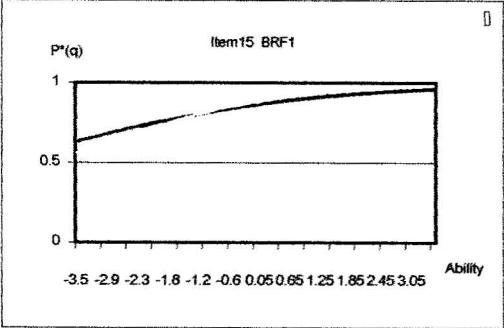
ขอบเขตลำดับขั้นที่ 4



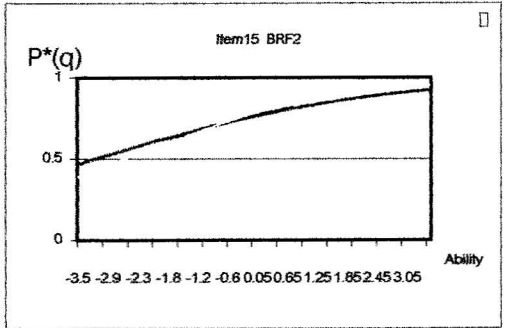
ขอบเขตลำดับขั้นที่ 5



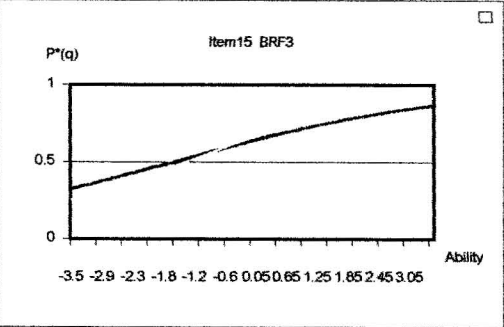
ข้อ 15 ขอบเขตลำดับขั้นที่ 1



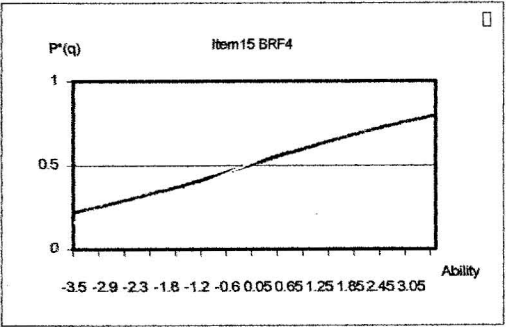
ขอบเขตลำดับขั้นที่ 2



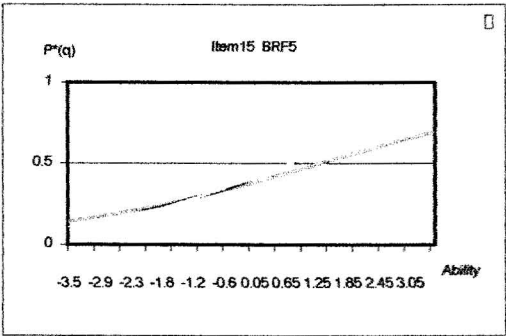
ขอบเขตลำดับขั้นที่ 3



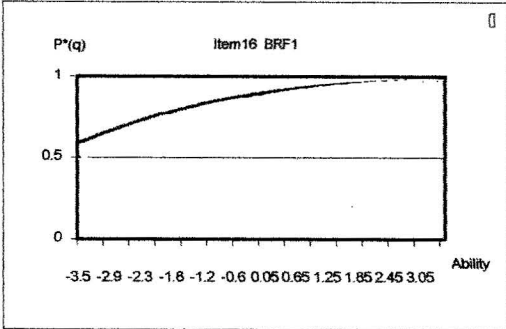
ขอบเขตลำดับขั้นที่ 4



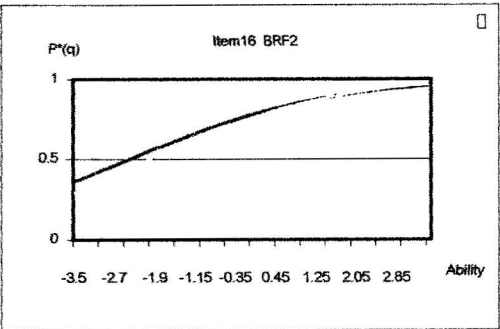
ขอบเขตลำดับขั้นที่ 5



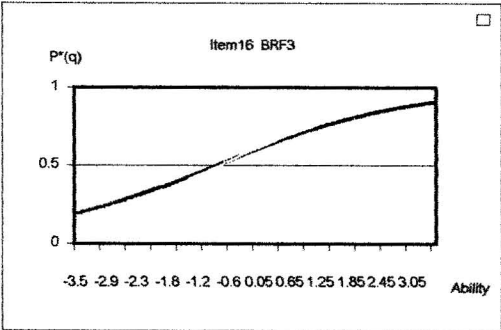
ข้อ 16 ขอบเขตลำดับขั้นที่ 1



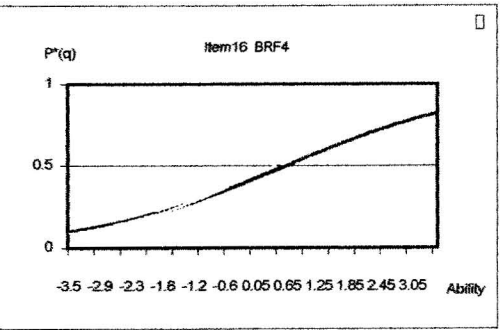
ขอบเขตลำดับขั้นที่ 2



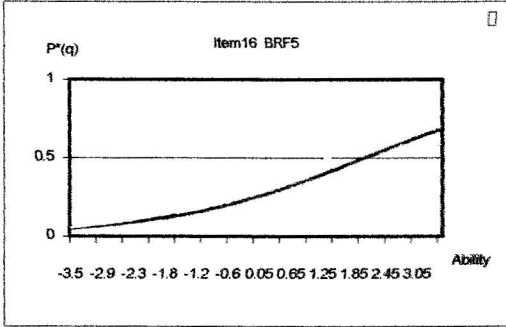
ขอบเขตลำดับขั้นที่ 3



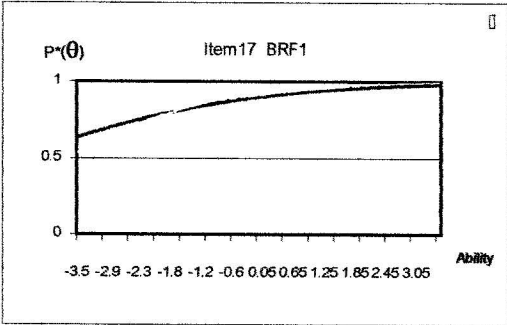
ขอบเขตลำดับขั้นที่ 4



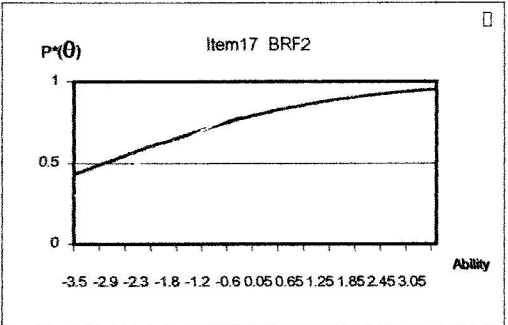
ขอบเขตลำดับขั้นที่ 5



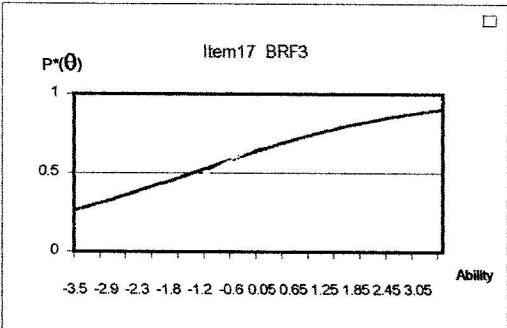
ข้อ 17 ขอบเขตลำดับขั้นที่ 1



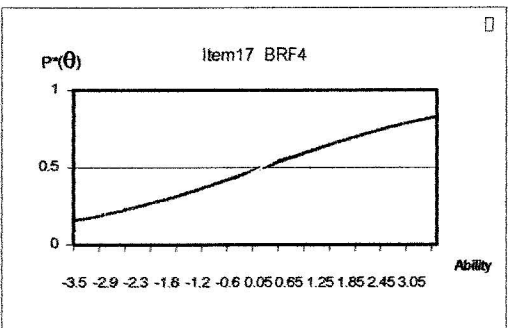
ขอบเขตลำดับขั้นที่ 2



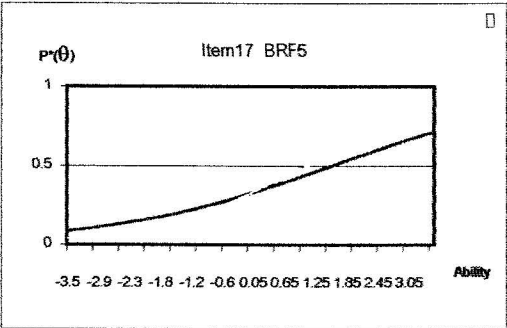
ขอบเขตลำดับขั้นที่ 3



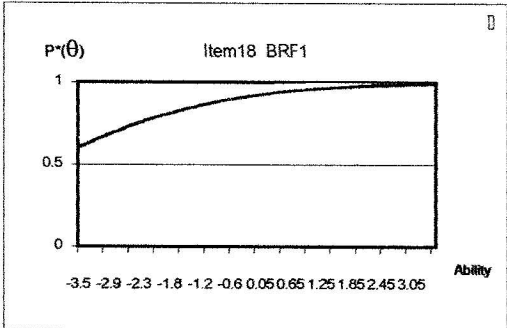
ขอบเขตลำดับขั้นที่ 4



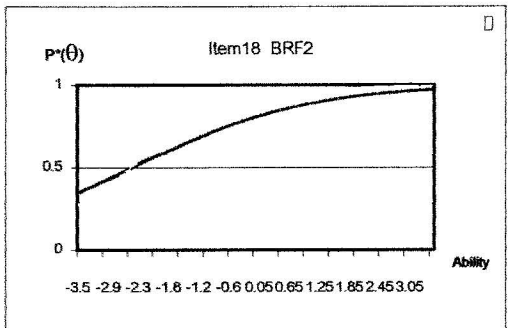
ขอบเขตลำดับขั้นที่ 5



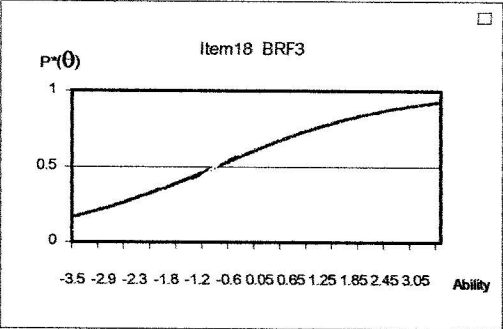
ข้อ 18 ขอบเขตลำดับขั้นที่ 1



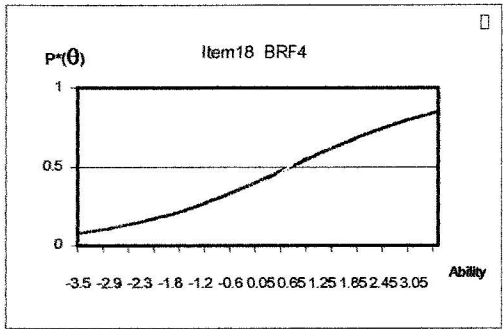
ขอบเขตลำดับขั้นที่ 2



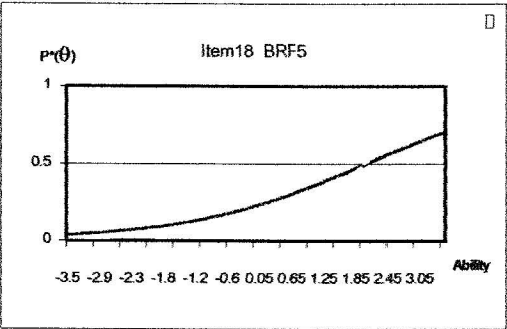
ขอบเขตลำดับขั้นที่ 3



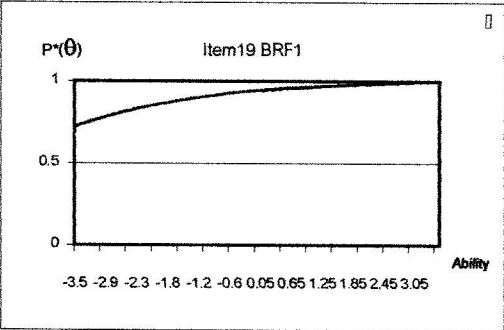
ขอบเขตลำดับขั้นที่ 4



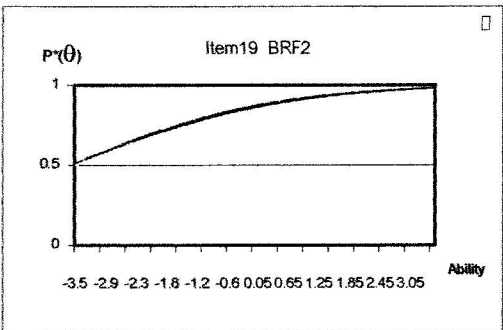
ขอบเขตลำดับขั้นที่ 5



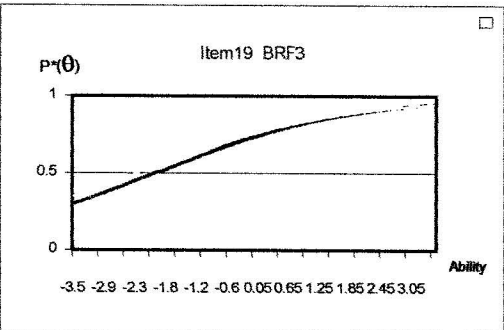
ข้อ 19 ขอบเขตลำดับขั้นที่ 1



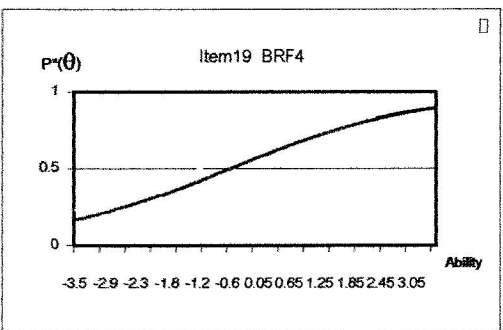
ขอบเขตลำดับขั้นที่ 2



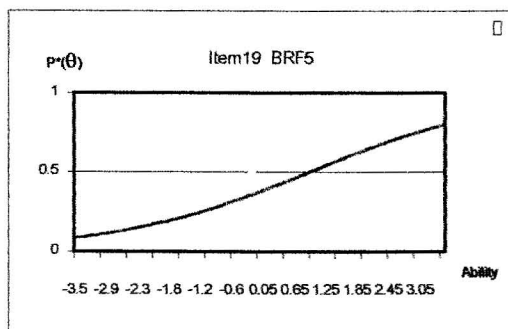
ขอบเขตลำดับขั้นที่ 3



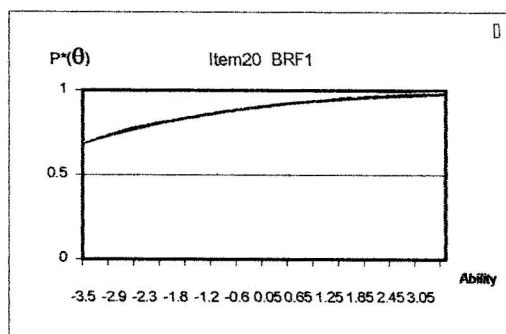
ขอบเขตลำดับขั้นที่ 4



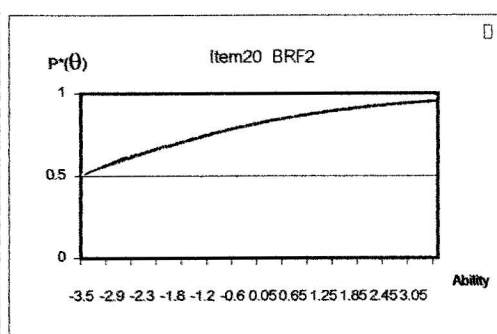
ขอบเขตลำดับขั้นที่ 5



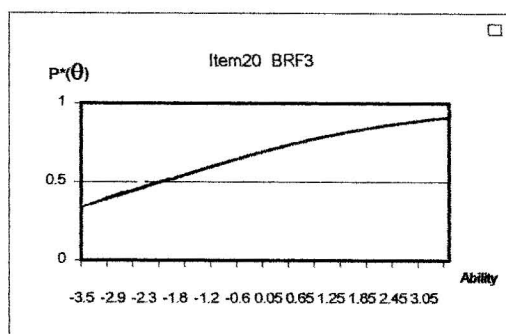
ข้อ 20 ขอบเขตลำดับขั้นที่ 1



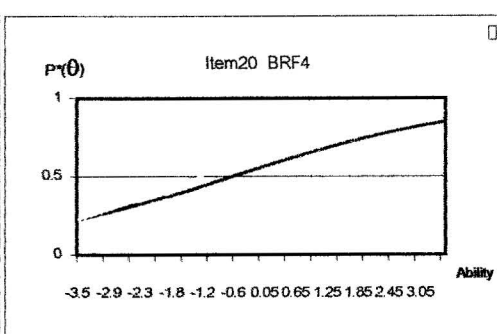
ขอบเขตลำดับขั้นที่ 2



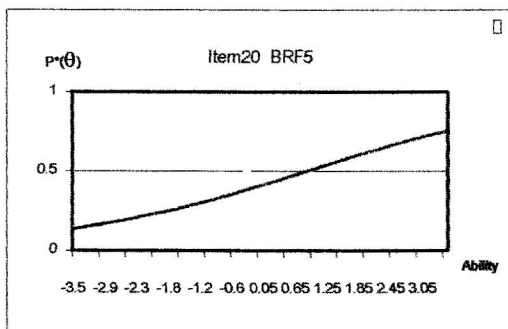
ขอบเขตลำดับขั้นที่ 3



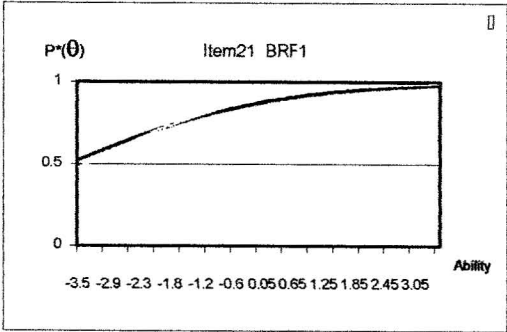
ขอบเขตลำดับขั้นที่ 4



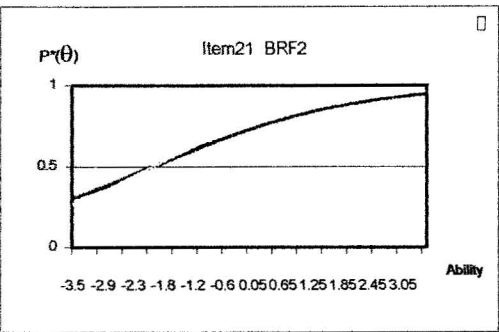
ขอบเขตลำดับขั้นที่ 5



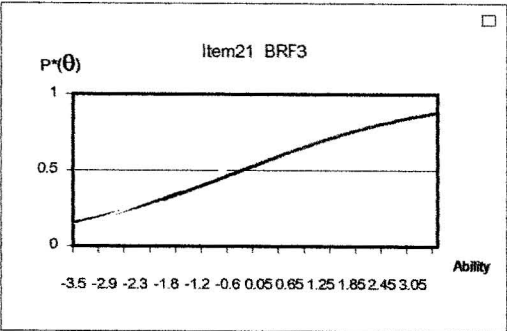
ข้อ 21 ขอบเขตลำดับขั้นที่ 1



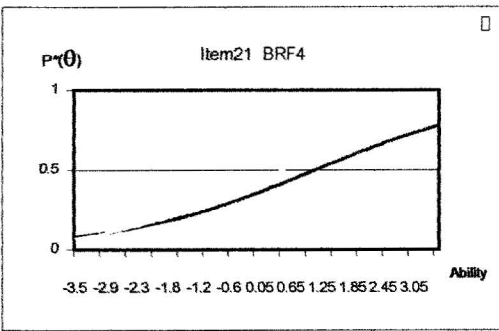
ข้อ 21 ขอบเขตลำดับขั้นที่ 2



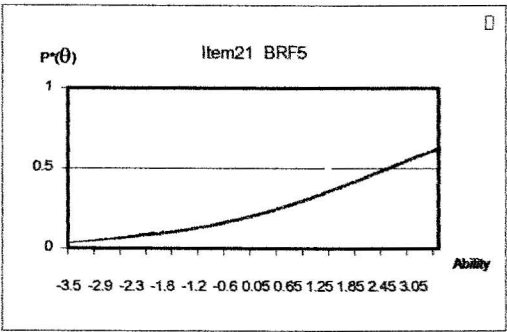
ข้อ 21 ขอบเขตลำดับขั้นที่ 3



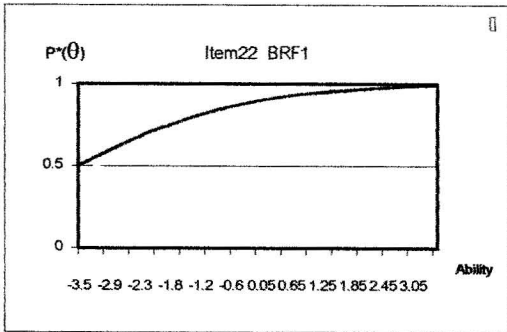
ข้อ 21 ขอบเขตลำดับขั้นที่ 4



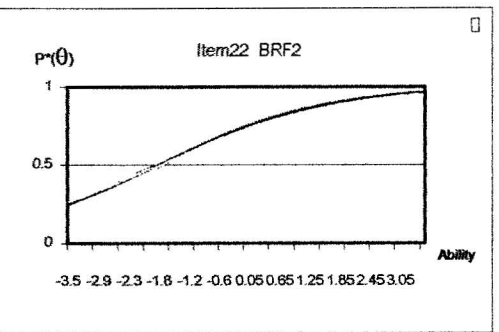
ข้อ 21 ขอบเขตลำดับขั้นที่ 5



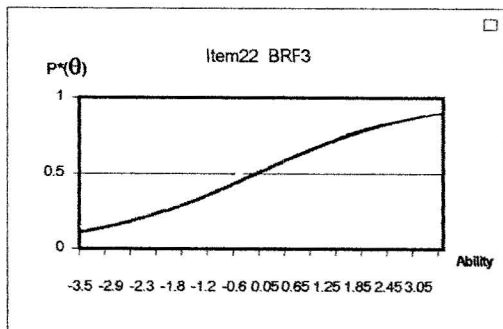
ข้อ 22 ขอบเขตลำดับขั้นที่ 1



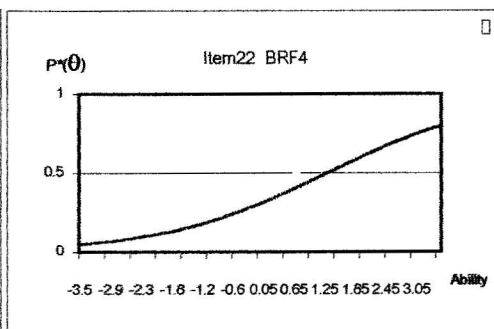
ข้อ 22 ขอบเขตลำดับขั้นที่ 2



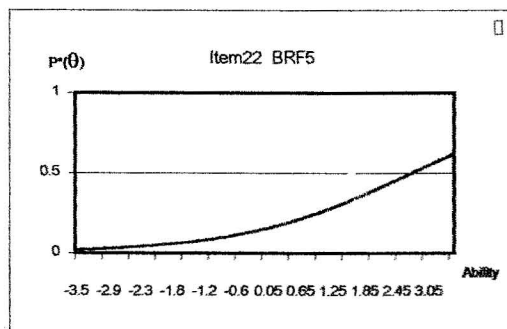
ขอบเขตลำดับขั้นที่ 3



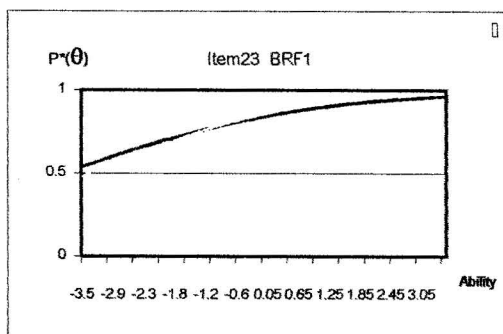
ขอบเขตลำดับขั้นที่ 4



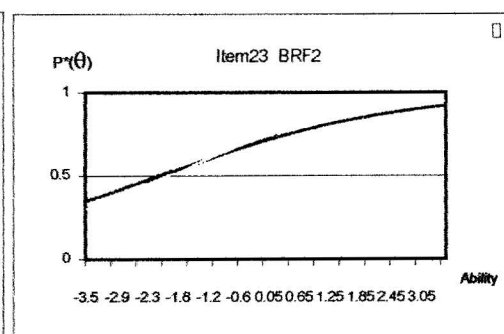
ขอบเขตลำดับขั้นที่ 5



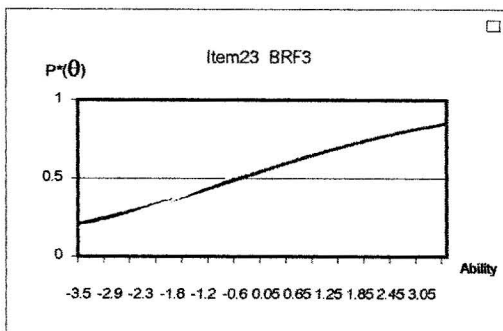
ข้อ 23 ขอบเขตลำดับขั้นที่ 1



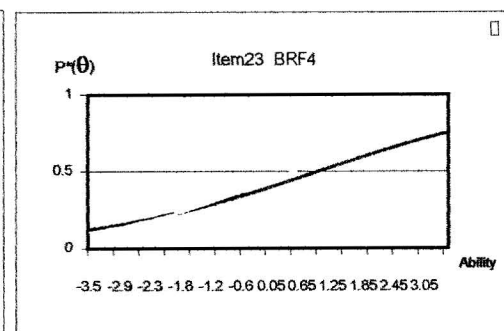
ขอบเขตลำดับขั้นที่ 2



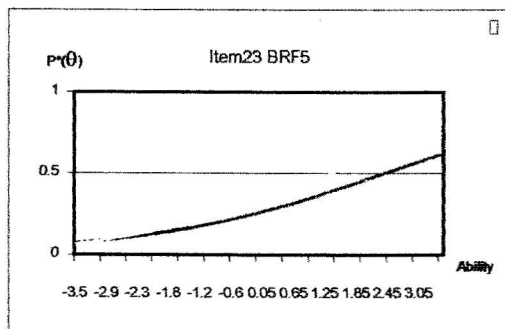
ขอบเขตลำดับขั้นที่ 3



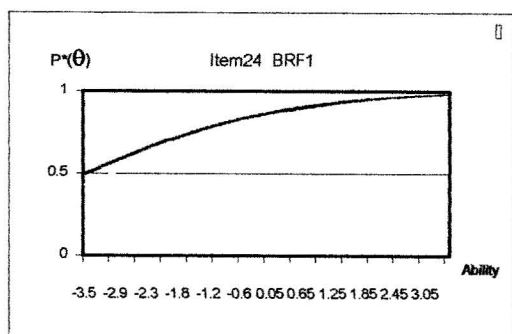
ขอบเขตลำดับขั้นที่ 4



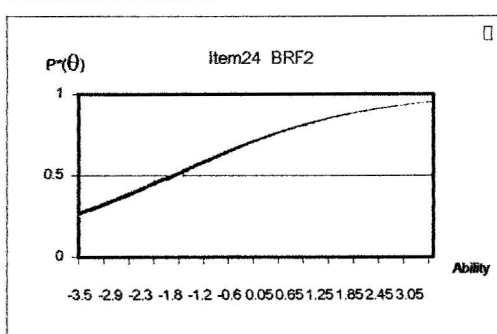
ขอบเขตลำดับขั้นที่ 5



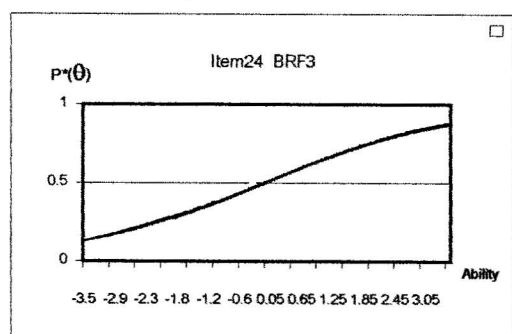
ข้อ 24 ขอบเขตลำดับขั้นที่ 1



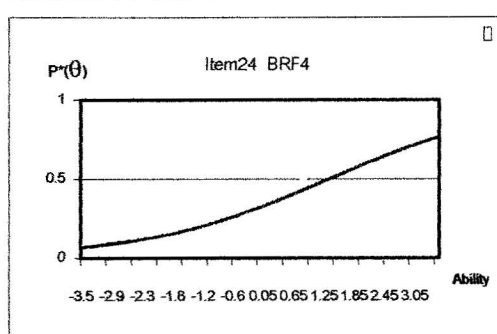
ขอบเขตลำดับขั้นที่ 2



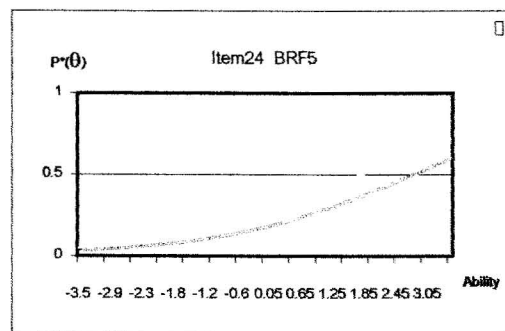
ขอบเขตลำดับขั้นที่ 3



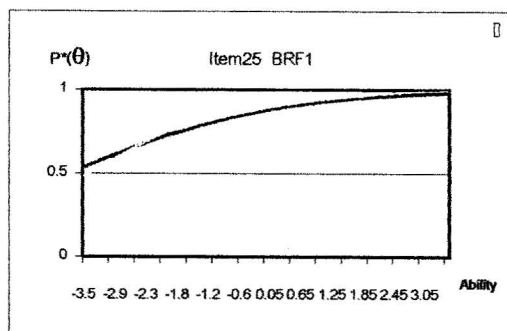
ขอบเขตลำดับขั้นที่ 4



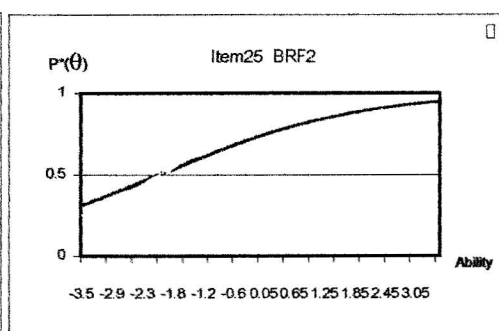
ขอบเขตลำดับขั้นที่ 5



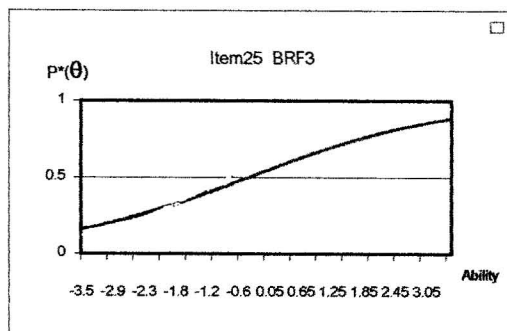
ข้อ 25 ขอบเขตลำดับขั้นที่ 1



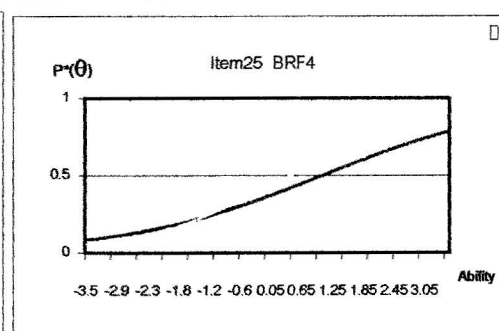
ขอบเขตลำดับขั้นที่ 2



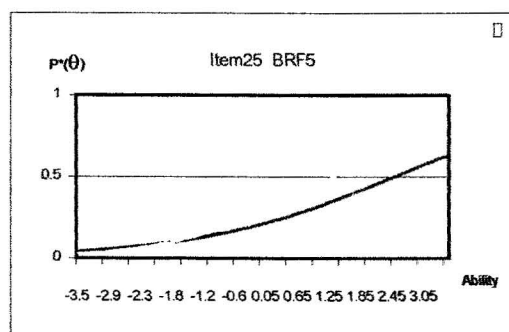
ขอบเขตลำดับขั้นที่ 3



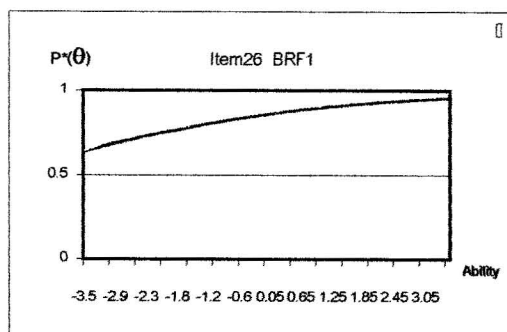
ขอบเขตลำดับขั้นที่ 4



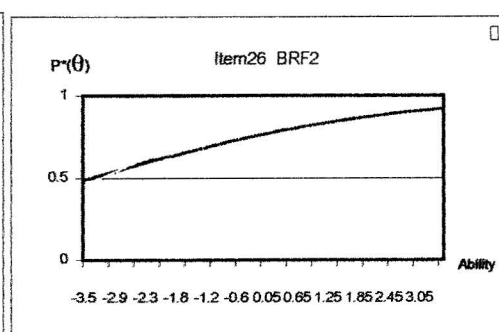
ขอบเขตลำดับขั้นที่ 5



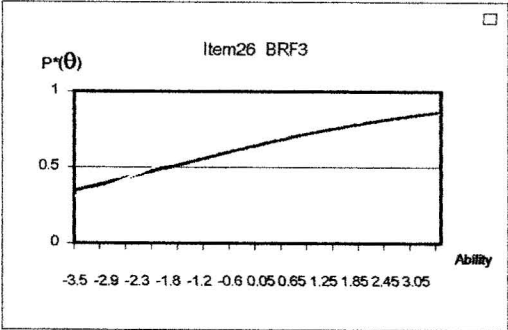
ข้อ 26 ขอบเขตลำดับขั้นที่ 1



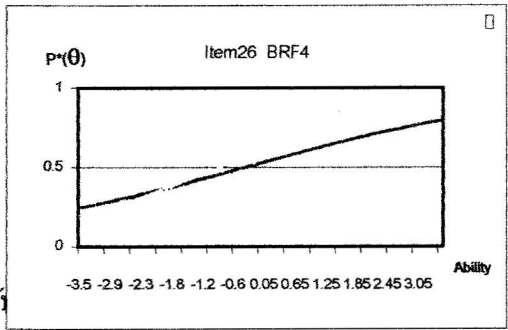
ขอบเขตลำดับขั้นที่ 2



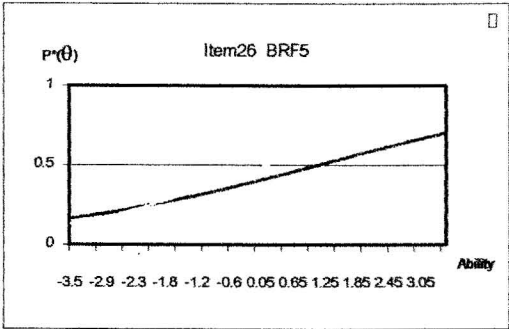
ขอบเขตลำดับขั้นที่ 3



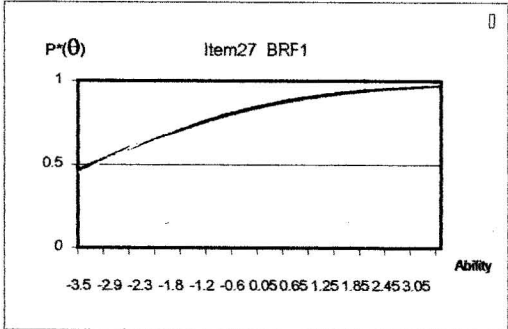
ขอบเขตลำดับขั้นที่ 4



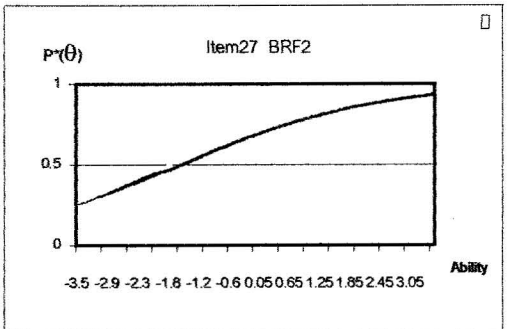
ขอบเขตลำดับขั้นที่ 5



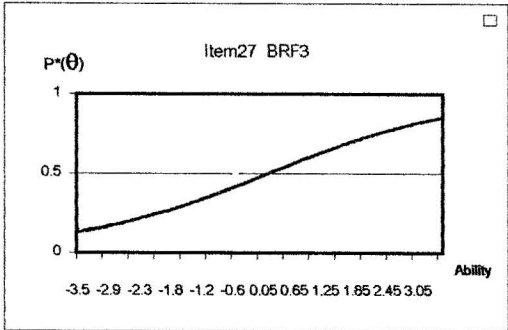
ข้อ 27 ขอบเขตลำดับขั้นที่ 1



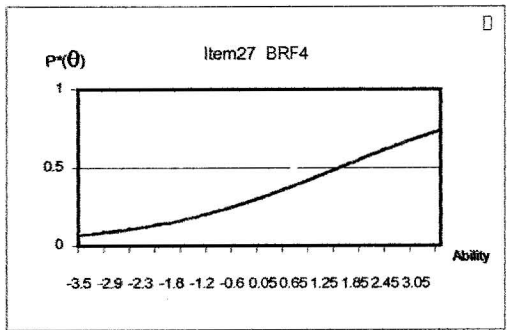
ขอบเขตลำดับขั้นที่ 2



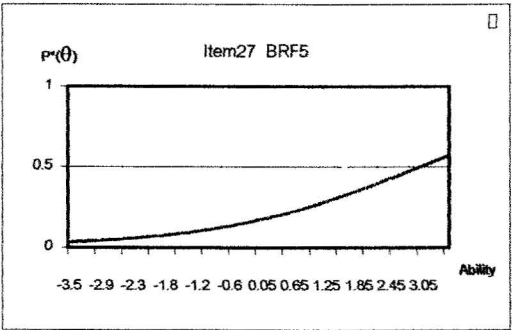
ขอบเขตลำดับขั้นที่ 3



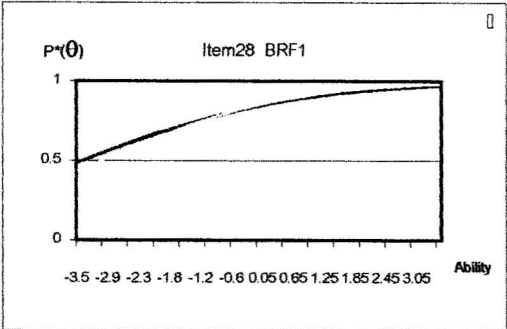
ขอบเขตลำดับขั้นที่ 4



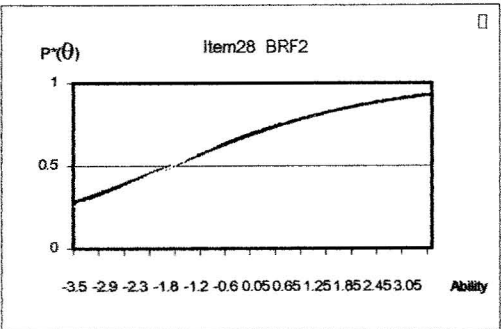
ขอบเขตลำดับขั้นที่ 5



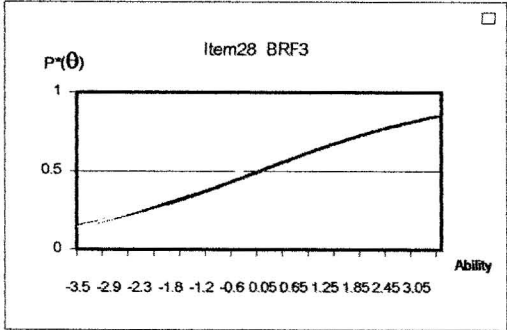
ข้อ 26 ขอบเขตลำดับขั้นที่ 1



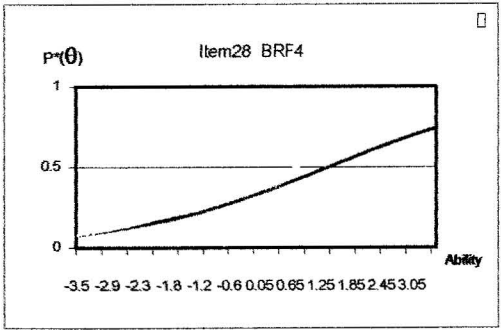
ขอบเขตลำดับขั้นที่ 2



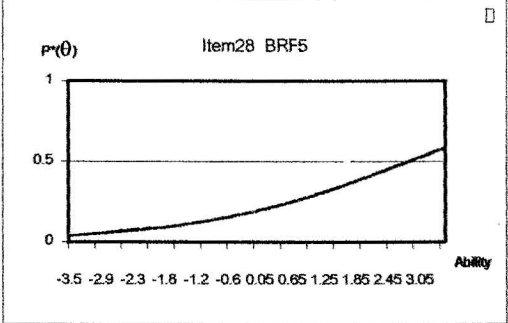
ขอบเขตลำดับขั้นที่ 3



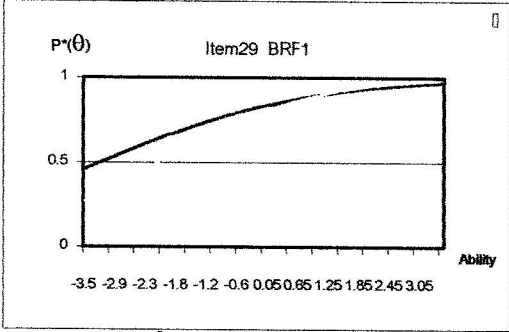
ขอบเขตลำดับขั้นที่ 4



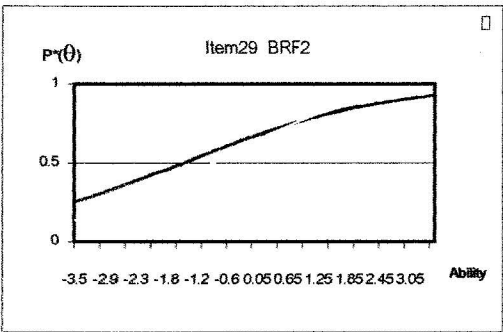
ขอบเขตลำดับขั้นที่ 5



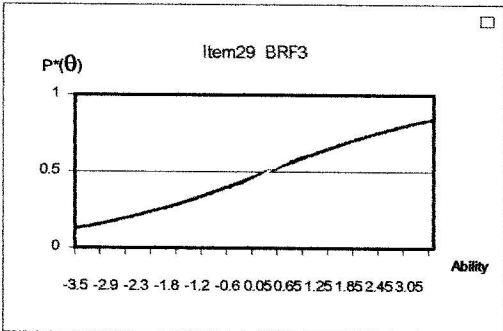
ข้อ 27 ขอบเขตลำดับขั้นที่ 1



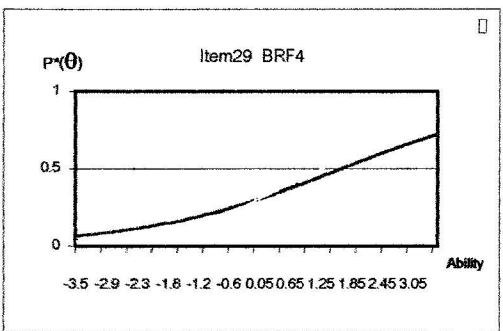
ข้อ 27 ขอบเขตลำดับขั้นที่ 2



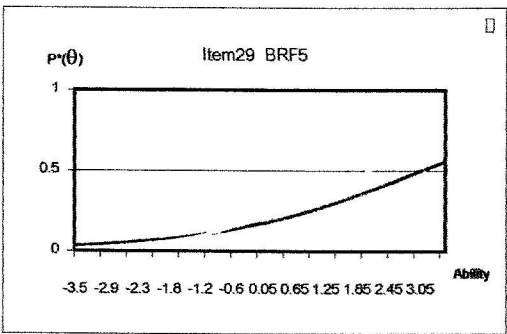
ข้อ 27 ขอบเขตลำดับขั้นที่ 3



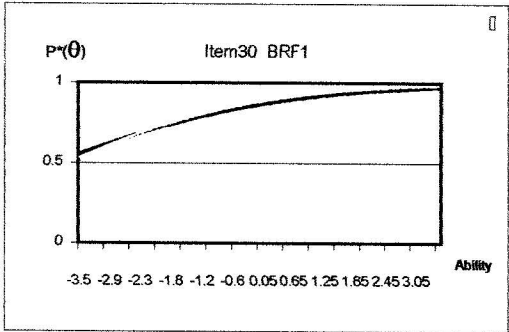
ข้อ 27 ขอบเขตลำดับขั้นที่ 4



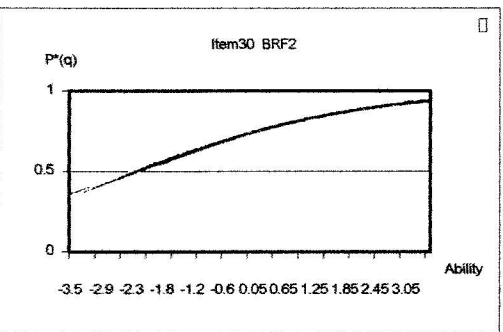
ข้อ 27 ขอบเขตลำดับขั้นที่ 5



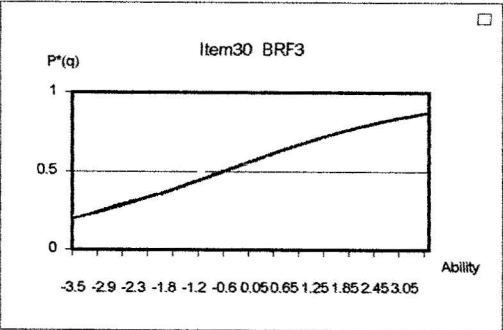
ข้อ 30 ขอบเขตลำดับขั้นที่ 1



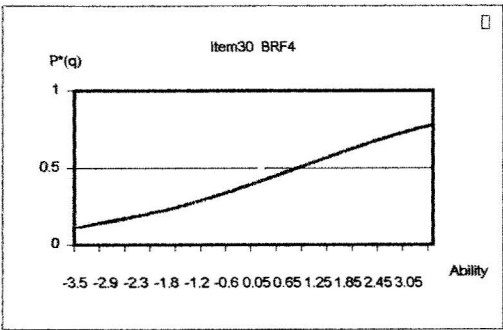
ข้อ 30 ขอบเขตลำดับขั้นที่ 2



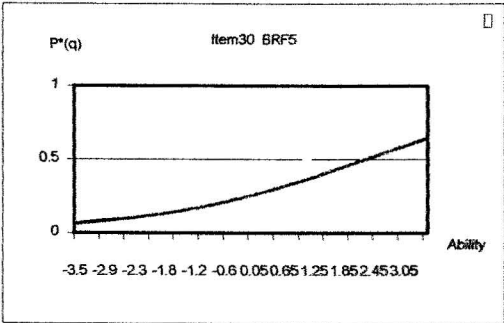
ขอบเขตลำดับขั้นที่ 3



ขอบเขตลำดับขั้นที่ 4



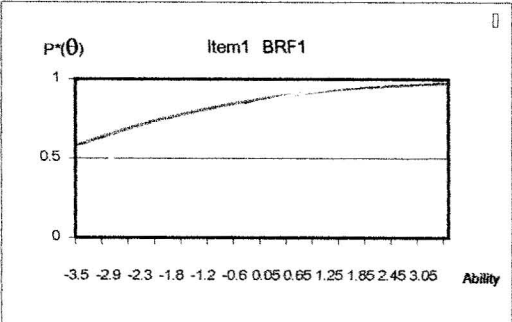
ขอบเขตลำดับขั้นที่ 5



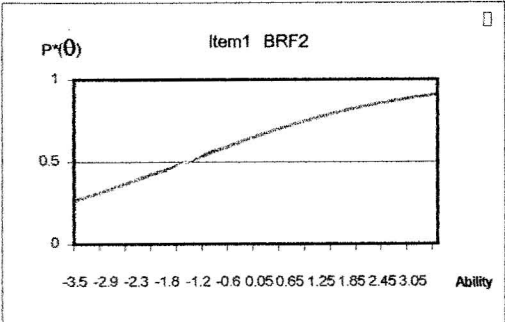
ฟังก์ชันผลการตอบข้อสอบในขอบเขตลำดับชั้น (BRF) ที่วิเคราะห์ด้วยวิธี GPCM ในแบบทดสอบขนาด 20 ข้อ
เมื่อกลุ่มตัวอย่างเพศชายต่อเพศหญิง เท่ากับ 1,000 คน ต่อ 1,000 คน

ชาย หญิง

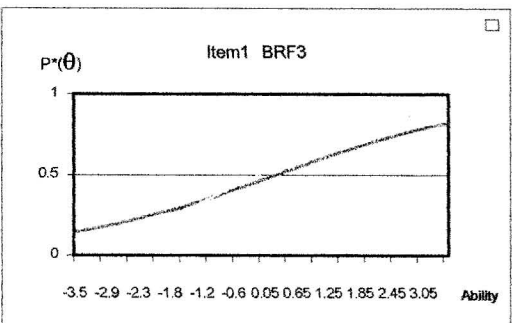
ข้อ 1 ขอบเขตลำดับชั้นที่ 1



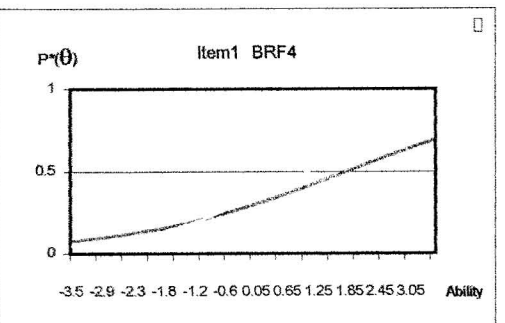
ขอบเขตลำดับชั้นที่ 2



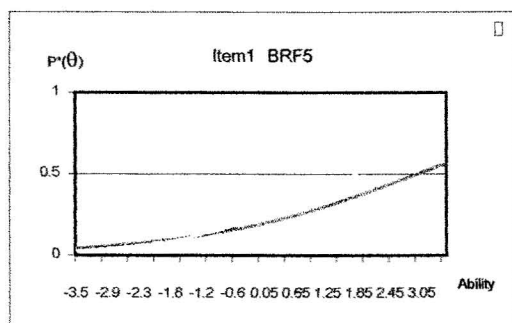
ขอบเขตลำดับชั้นที่ 3



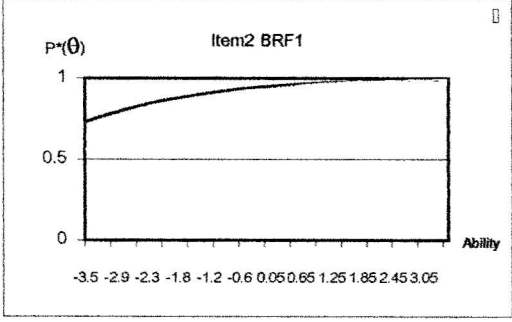
ขอบเขตลำดับชั้นที่ 4



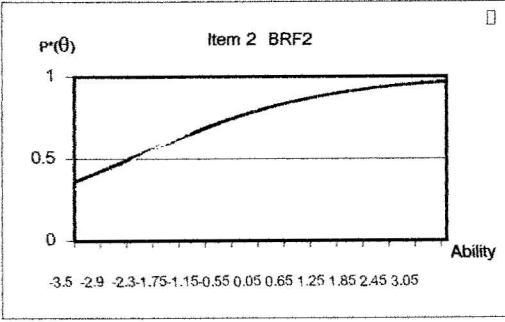
ขอบเขตลำดับชั้นที่ 5



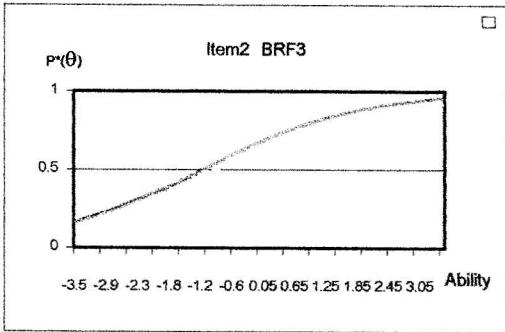
ข้อ 2 ขอบเขตลำดับชั้นที่ 1



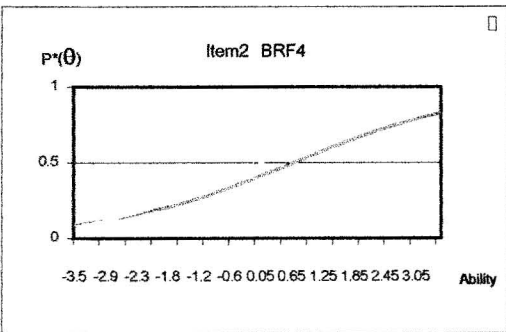
ขอบเขตลำดับชั้นที่ 2



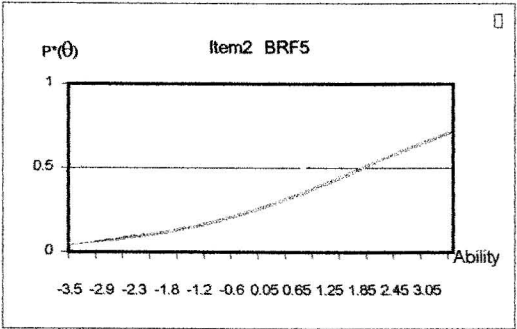
ขอบเขตลำดับขั้นที่ 3



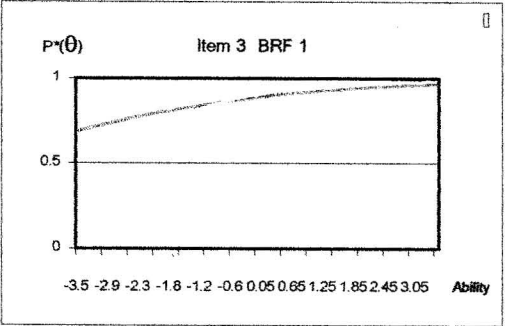
ขอบเขตลำดับขั้นที่ 4



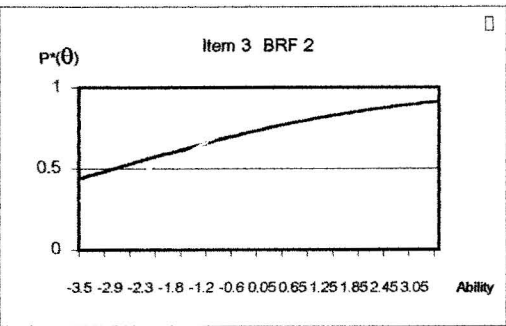
ขอบเขตลำดับขั้นที่ 5



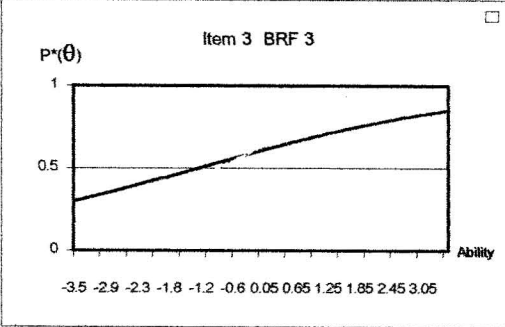
ข้อ 3 ขอบเขตลำดับขั้นที่ 1



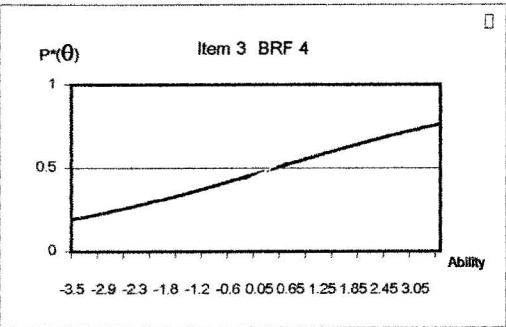
ขอบเขตลำดับขั้นที่ 2



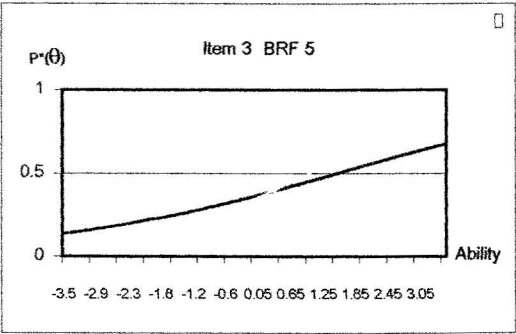
ขอบเขตลำดับขั้นที่ 3



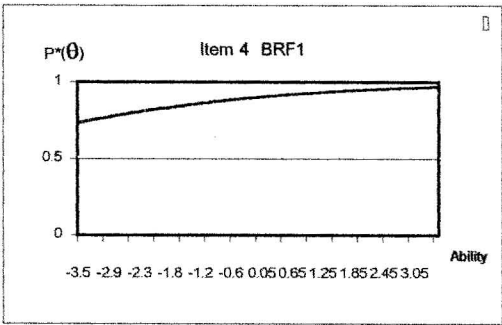
ขอบเขตลำดับขั้นที่ 4



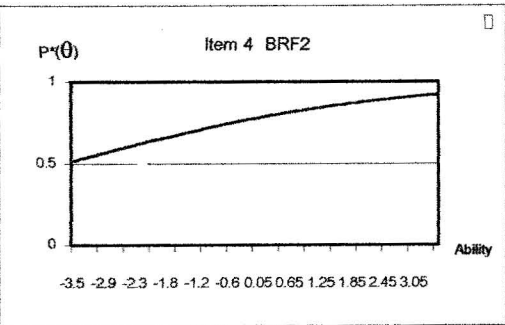
ขอเบตล่ำดบ้ันที่ 5



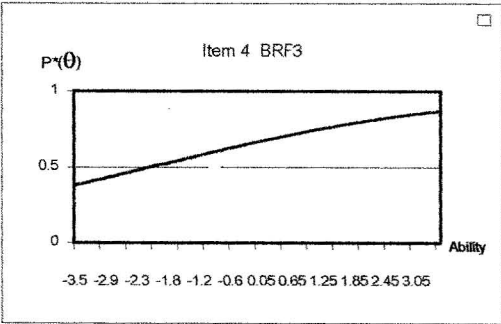
ข้อ 4 ขอเบตล่ำดบ้ันที่ 1



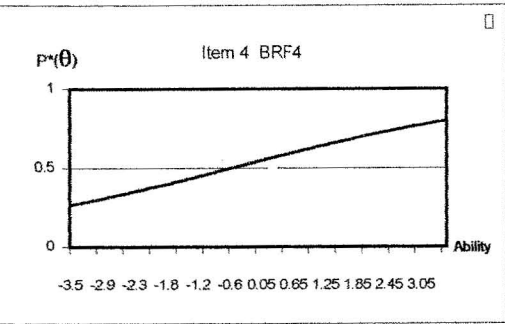
ขอเบตล่ำดบ้ันที่ 2



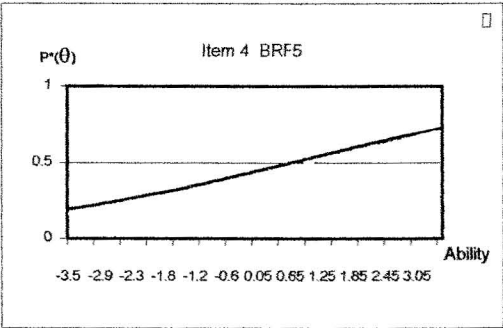
ขอเบตล่ำดบ้ันที่ 3



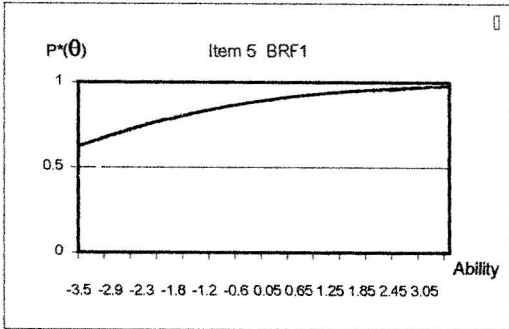
ขอเบตล่ำดบ้ันที่ 4



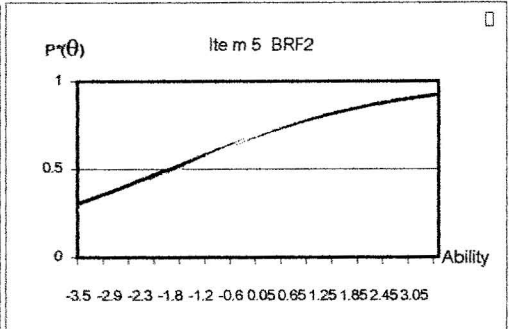
ขอเบตล่ำดบ้ันที่ 5



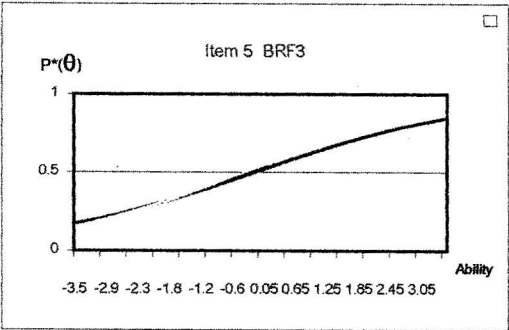
ข้อ 5 ขอบเขตลำดับขั้นที่ 1



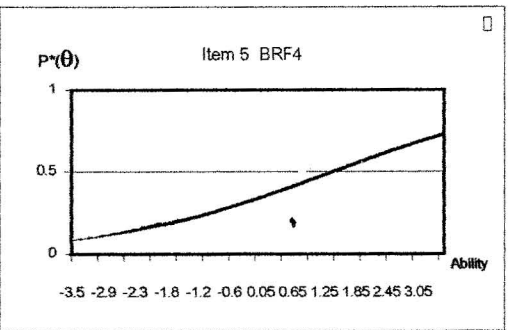
ขอบเขตลำดับขั้นที่ 2



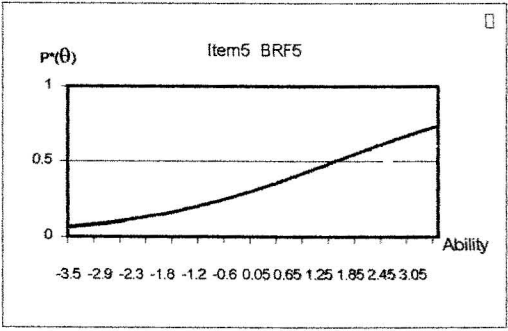
ขอบเขตลำดับขั้นที่ 3



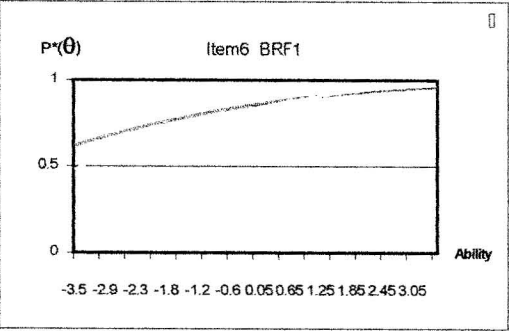
ขอบเขตลำดับขั้นที่ 4



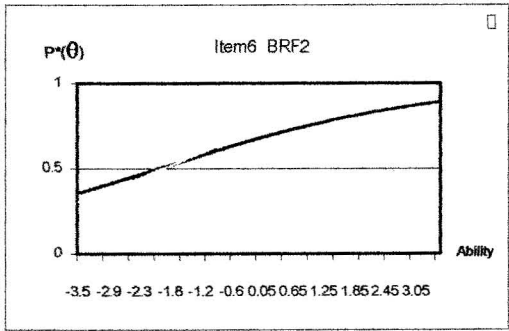
ขอบเขตลำดับขั้นที่ 5



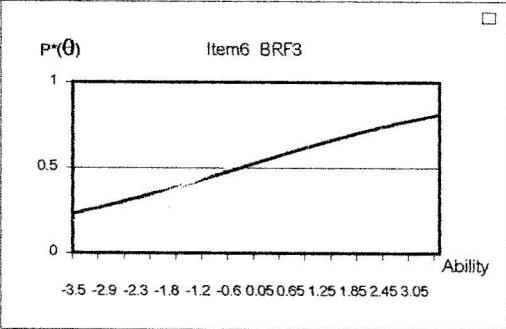
ข้อ 6 ขอบเขตลำดับขั้นที่ 1



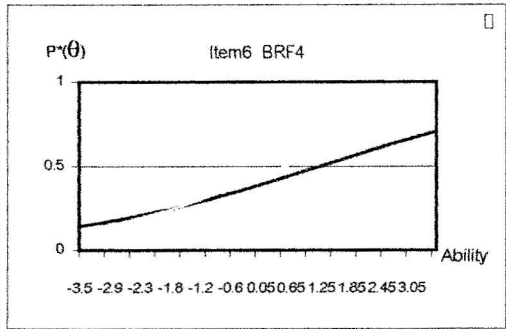
ขอบเขตลำดับขั้นที่ 2



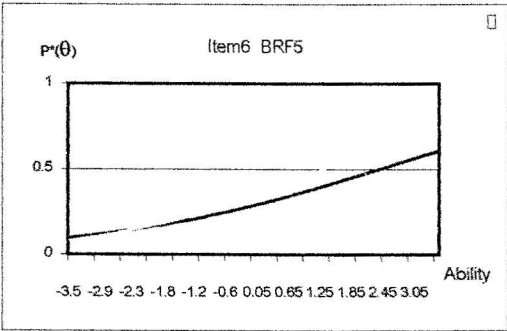
ขอบเขตลำดับขั้นที่ 3



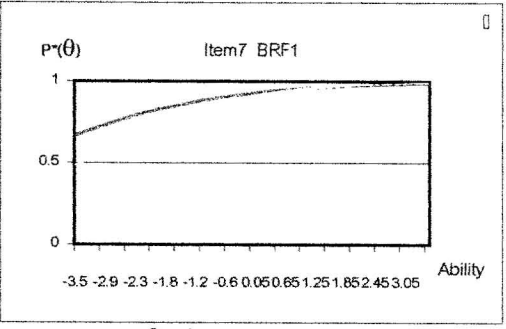
ขอบเขตลำดับขั้นที่ 4



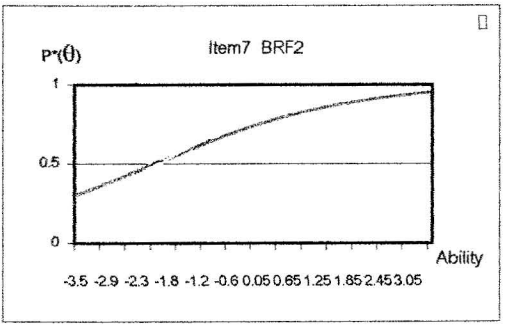
ขอบเขตลำดับขั้นที่ 5



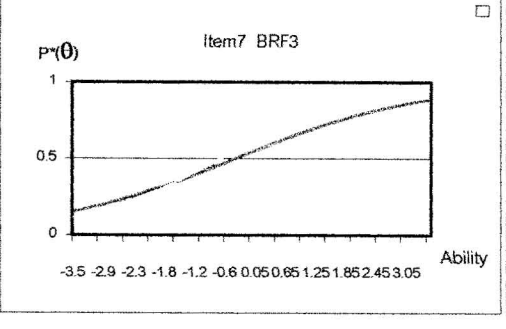
ข้อ 7 ขอบเขตลำดับขั้นที่ 1



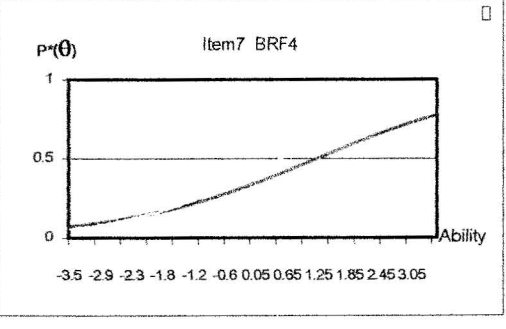
ขอบเขตลำดับขั้นที่ 2



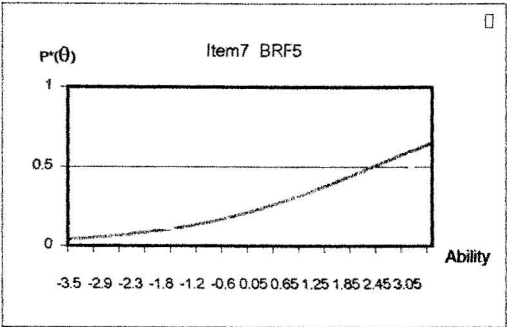
ขอบเขตลำดับขั้นที่ 3



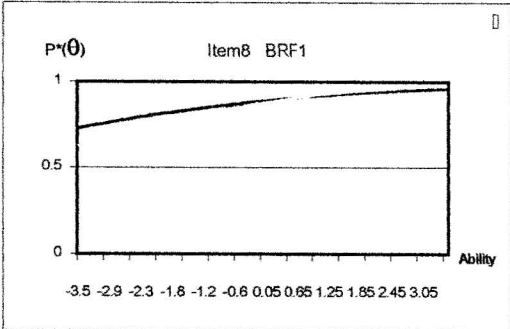
ขอบเขตลำดับขั้นที่ 4



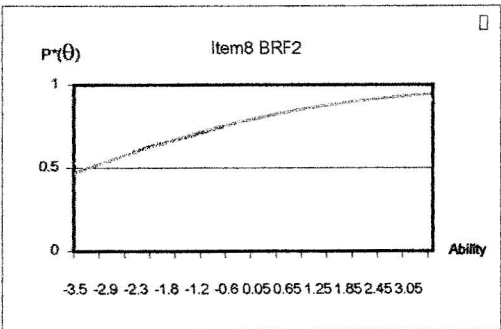
ข้อ 8 ขอบเขตลำดับขั้นที่ 5



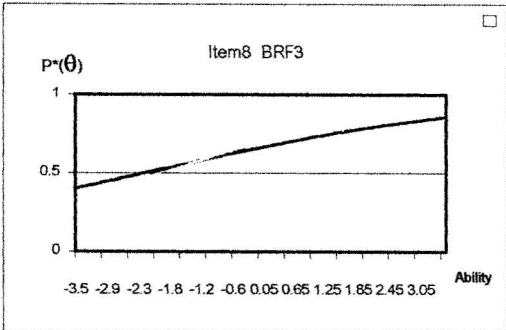
ข้อ 8 ขอบเขตลำดับขั้นที่ 1



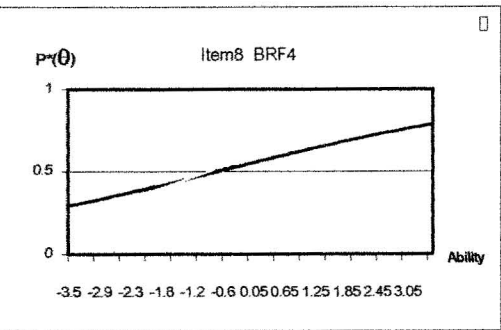
ข้อ 8 ขอบเขตลำดับขั้นที่ 2



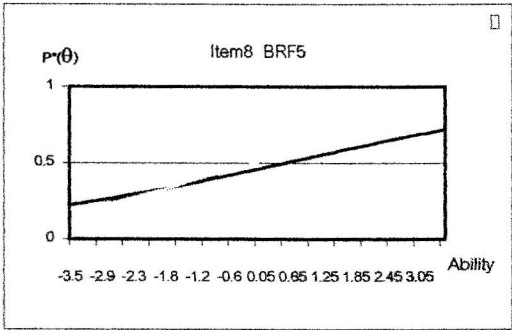
ข้อ 8 ขอบเขตลำดับขั้นที่ 3



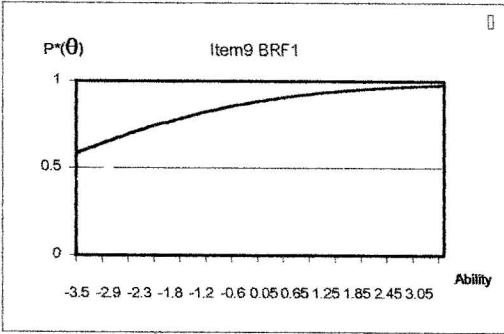
ข้อ 8 ขอบเขตลำดับขั้นที่ 4



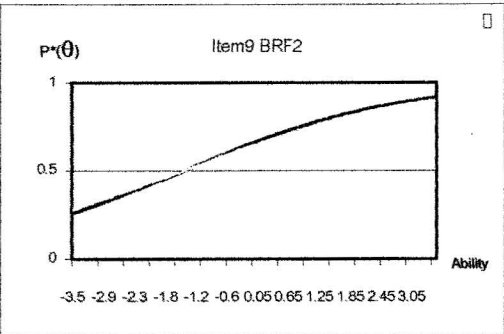
ข้อ 8 ขอบเขตลำดับขั้นที่ 5



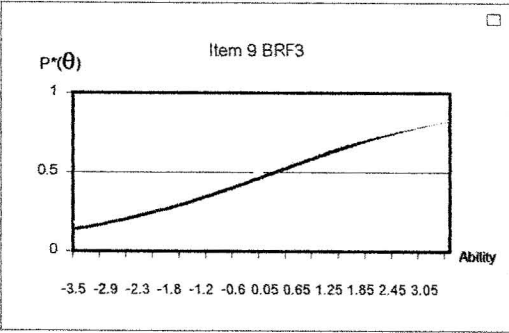
ข้อ 9 ขอบเขตลำดับขั้นที่ 1



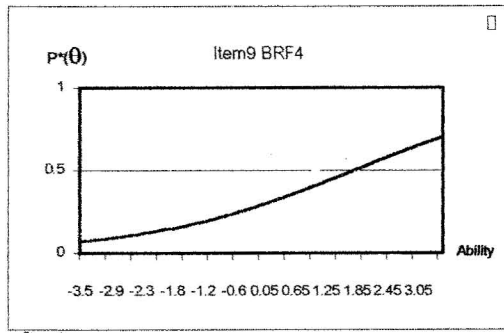
ข้อ 9 ขอบเขตลำดับขั้นที่ 2



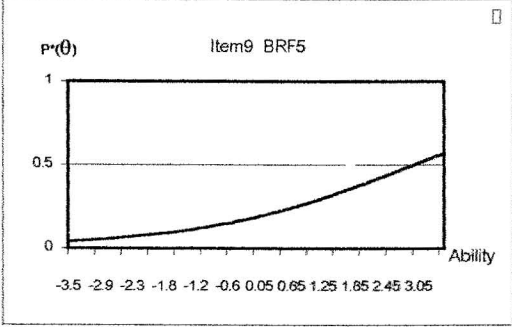
ข้อ 9 ขอบเขตลำดับขั้นที่ 3



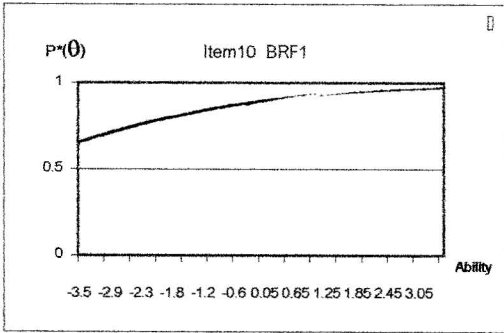
ข้อ 9 ขอบเขตลำดับขั้นที่ 4



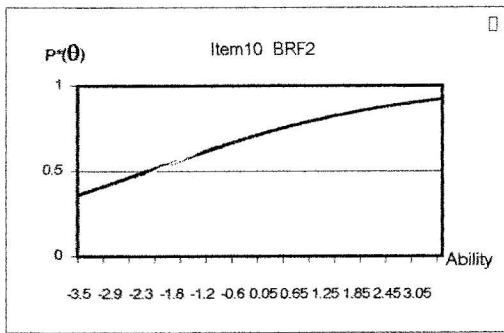
ข้อ 9 ขอบเขตลำดับขั้นที่ 5



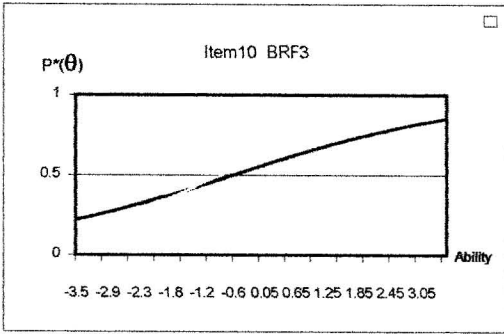
ข้อ 10 ขอบเขตลำดับขั้นที่ 1



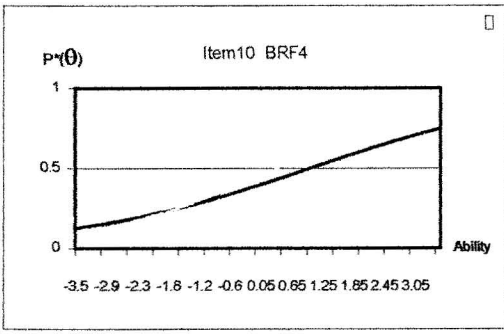
ข้อ 10 ขอบเขตลำดับขั้นที่ 2



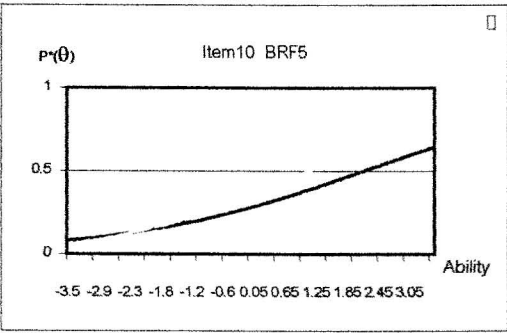
ขอบเขตลำดับขั้นที่ 3



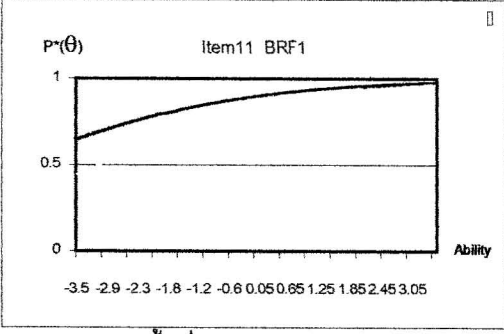
ขอบเขตลำดับขั้นที่ 4



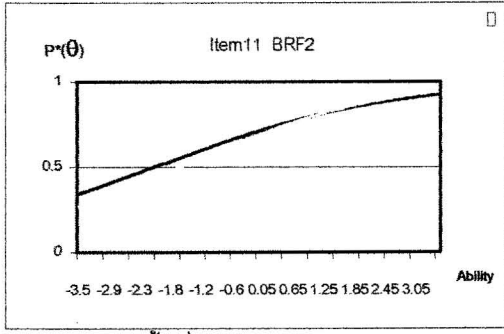
ขอบเขตลำดับขั้นที่ 5



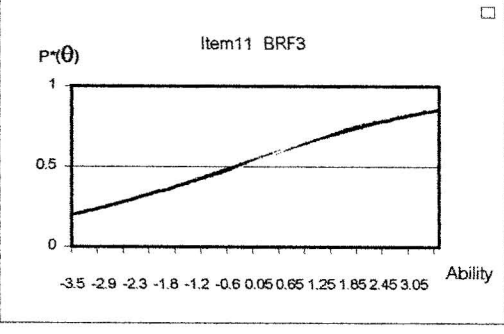
ข้อ 11 ขอบเขตลำดับขั้นที่ 1



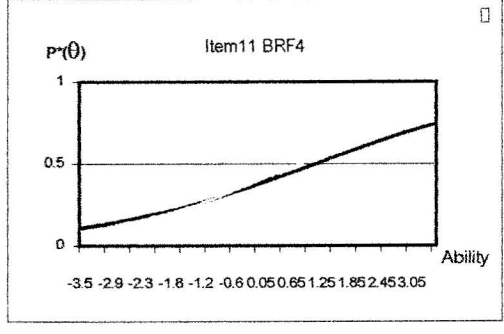
ขอบเขตลำดับขั้นที่ 2



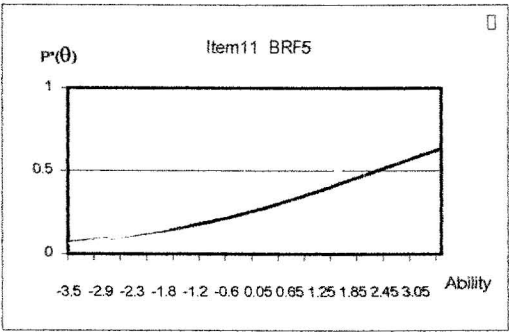
ขอบเขตลำดับขั้นที่ 3



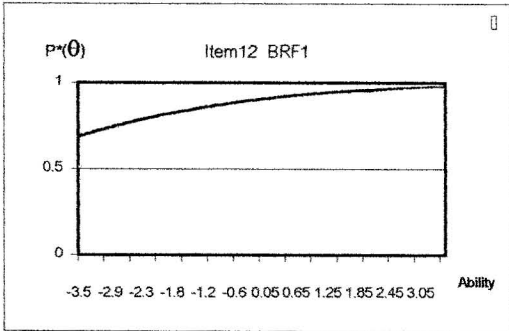
ขอบเขตลำดับขั้นที่ 4



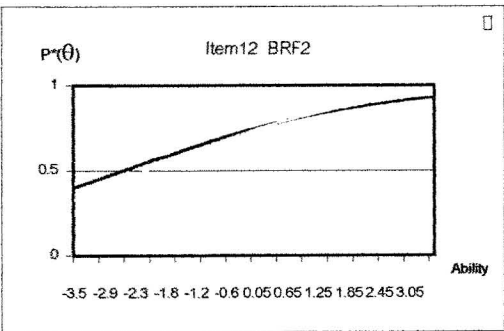
ขอบเขตลำดับขั้นที่ 5



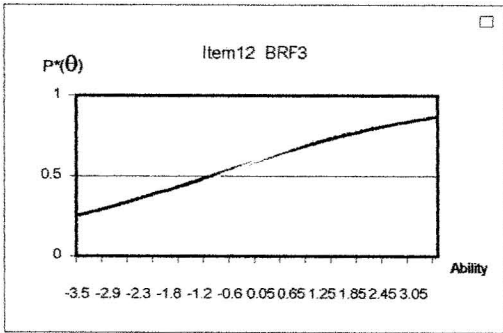
ข้อ 12 ขอบเขตลำดับขั้นที่ 1



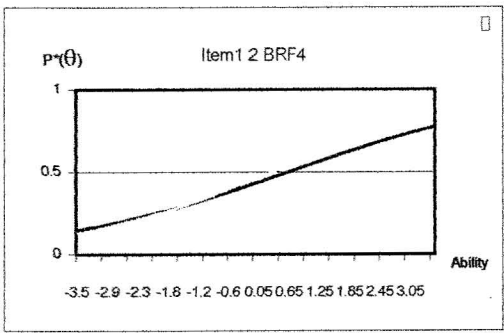
ขอบเขตลำดับขั้นที่ 2



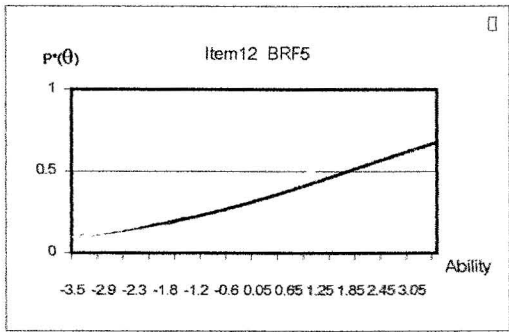
ขอบเขตลำดับขั้นที่ 3



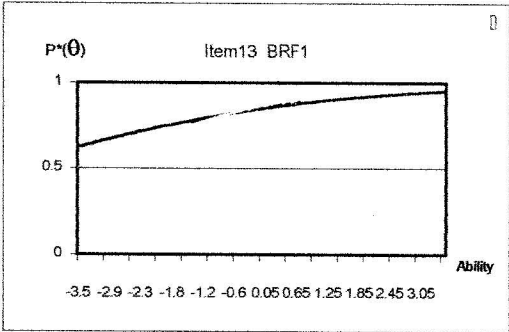
ขอบเขตลำดับขั้นที่ 4



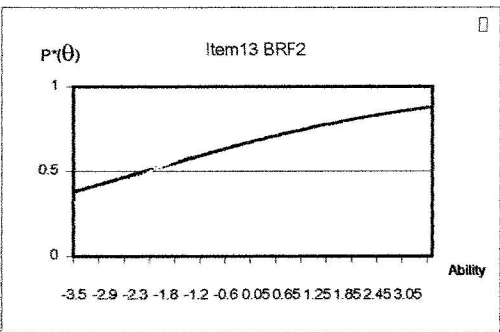
ขอบเขตลำดับขั้นที่ 5



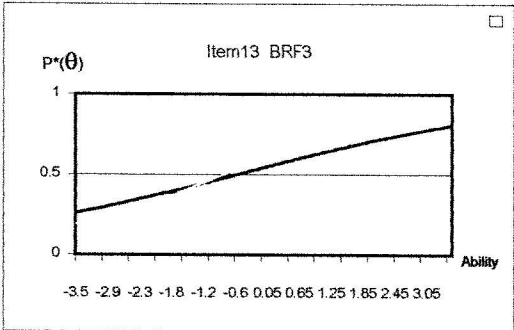
ข้อ 13 ขอบเขตลำดับขั้นที่ 1



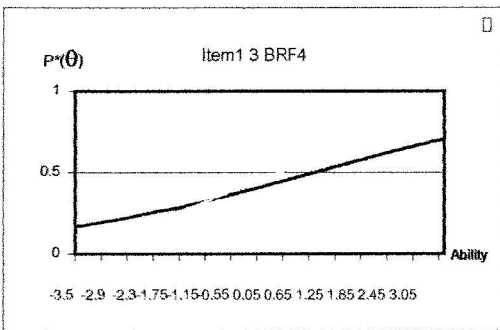
ขอบเขตลำดับขั้นที่ 2



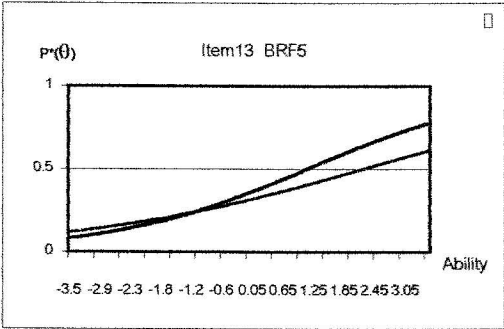
ขอบเขตลำดับขั้นที่ 3



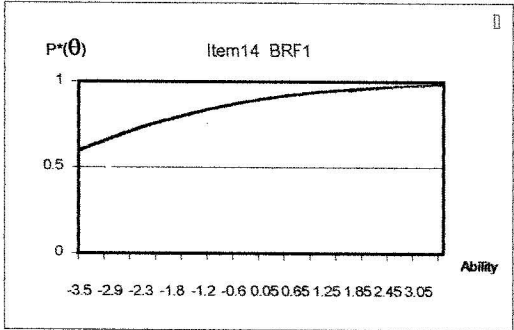
ขอบเขตลำดับขั้นที่ 4



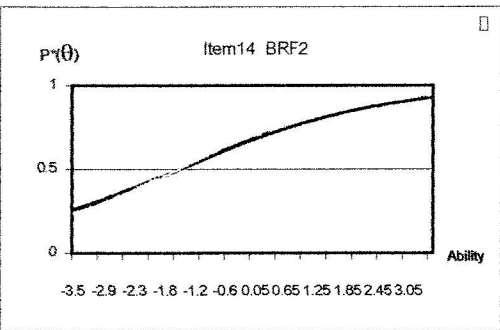
ขอบเขตลำดับขั้นที่ 5



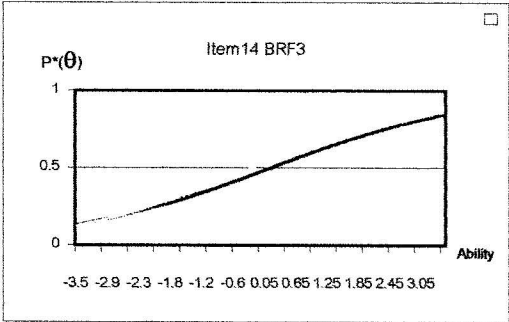
ข้อ 14 ขอบเขตลำดับขั้นที่ 1



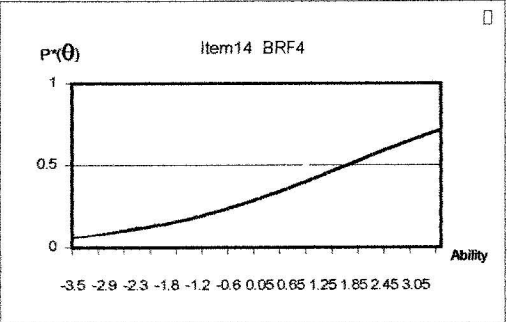
ขอบเขตลำดับขั้นที่ 2



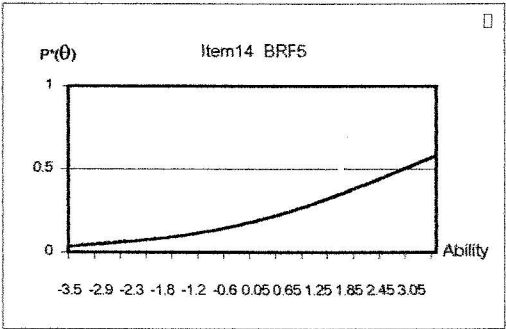
ข้อ 14 ขอบเขตลำดับขั้นที่ 3



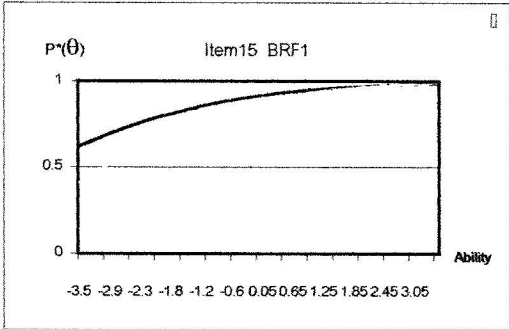
ข้อ 14 ขอบเขตลำดับขั้นที่ 4



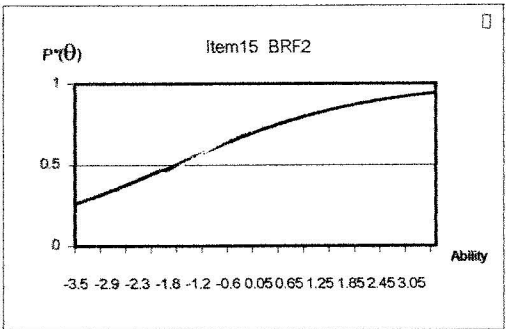
ข้อ 14 ขอบเขตลำดับขั้นที่ 5



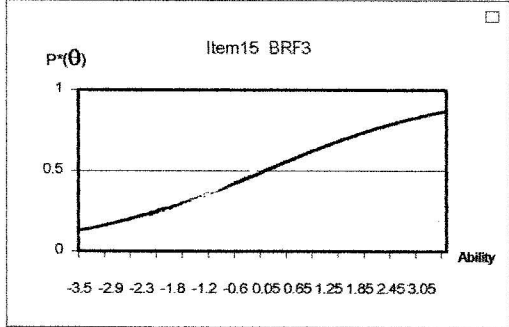
ข้อ 15 ขอบเขตลำดับขั้นที่ 1



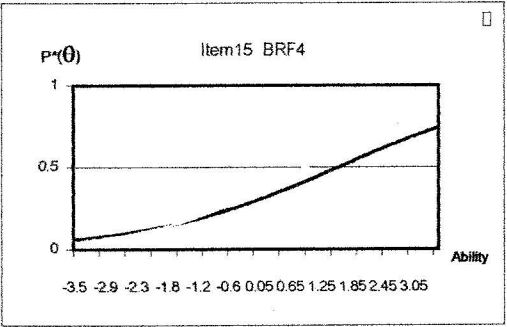
ข้อ 15 ขอบเขตลำดับขั้นที่ 2



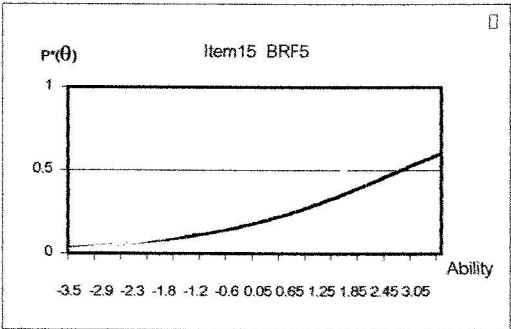
ข้อ 15 ขอบเขตลำดับขั้นที่ 3



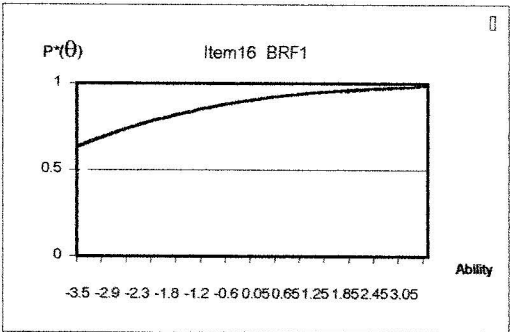
ข้อ 15 ขอบเขตลำดับขั้นที่ 4



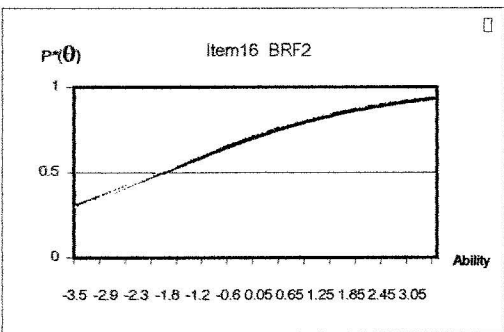
ขอบเขตลำดับขั้นที่ 5



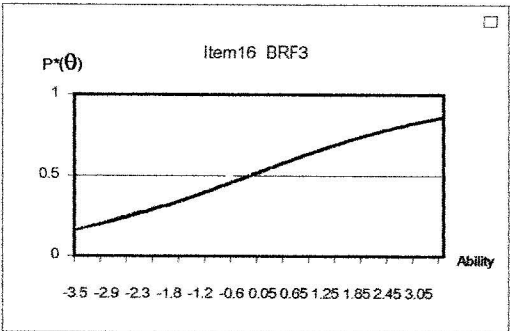
ข้อ 16 ขอบเขตลำดับขั้นที่ 1



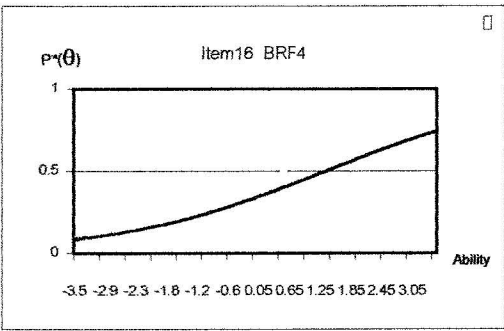
ขอบเขตลำดับขั้นที่ 2



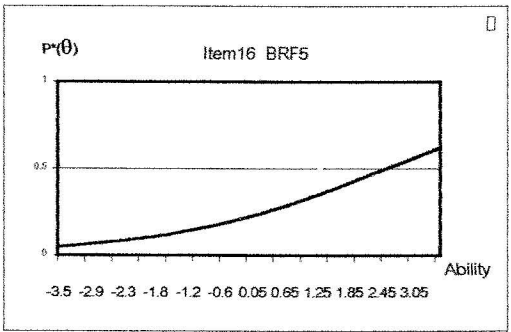
ขอบเขตลำดับขั้นที่ 3



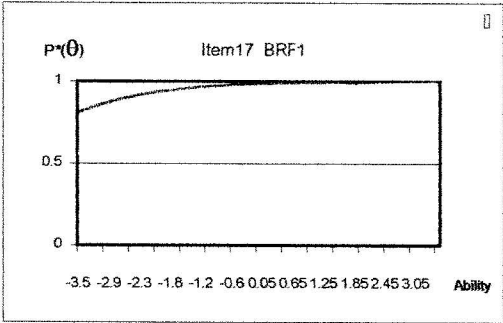
ขอบเขตลำดับขั้นที่ 4



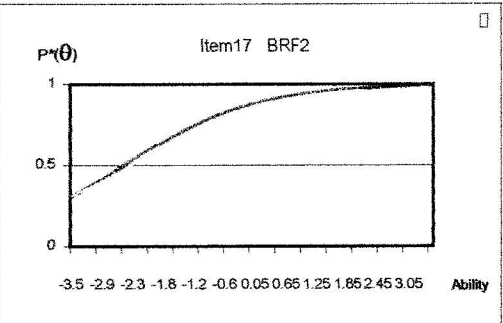
ขอบเขตลำดับขั้นที่ 5



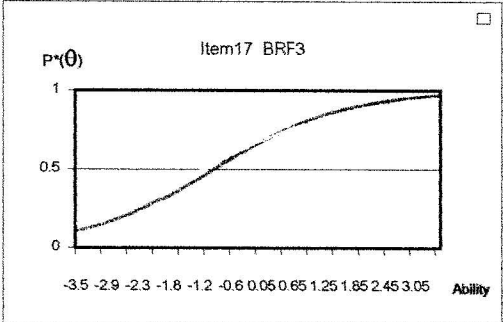
ข้อ 17 ขอบเขตลำดับขั้นที่ 1



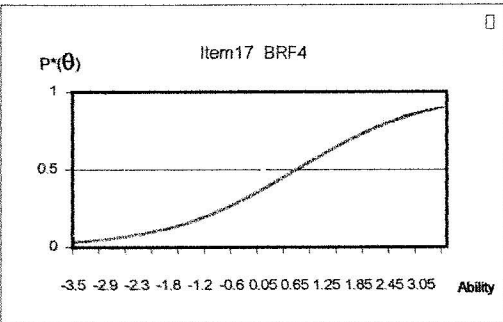
ข้อ 17 ขอบเขตลำดับขั้นที่ 2



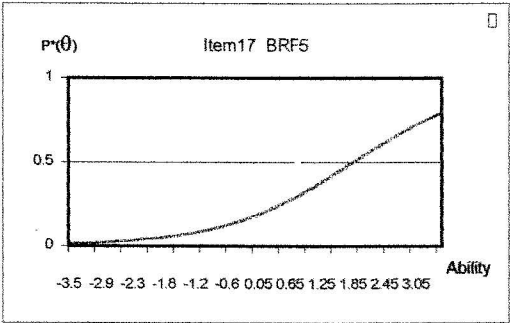
ข้อ 17 ขอบเขตลำดับขั้นที่ 3



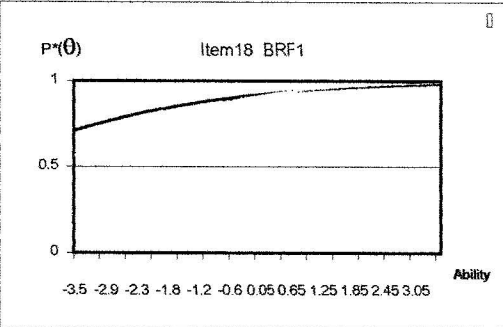
ข้อ 17 ขอบเขตลำดับขั้นที่ 4



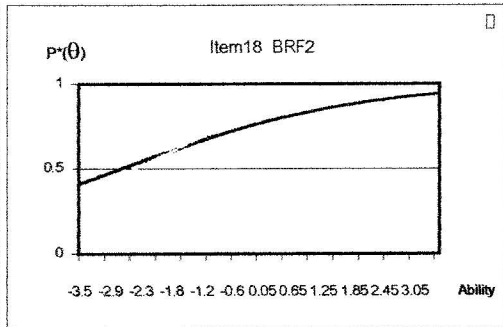
ข้อ 17 ขอบเขตลำดับขั้นที่ 5



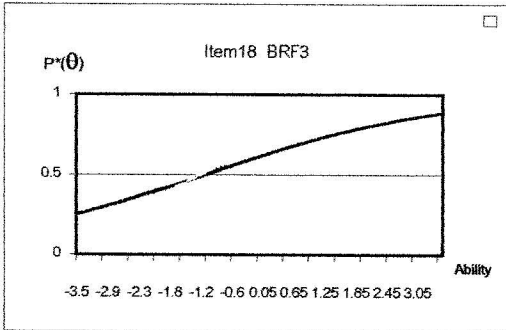
ข้อ 18 ขอบเขตลำดับขั้นที่ 1



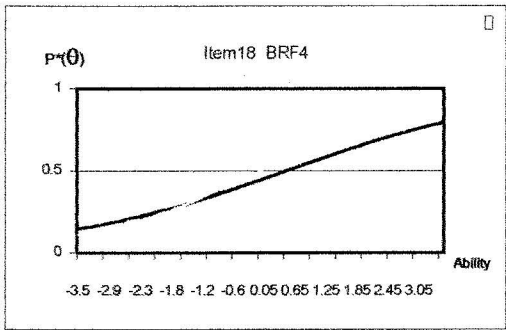
ข้อ 18 ขอบเขตลำดับขั้นที่ 2



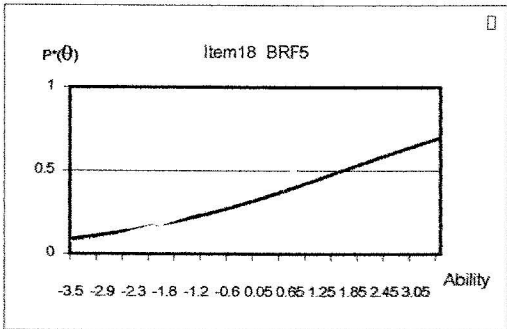
ขอบเขตลำดับขั้นที่ 3



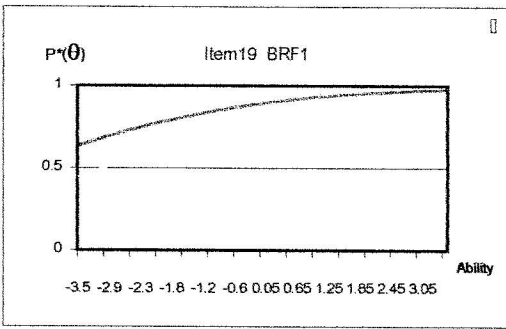
ขอบเขตลำดับขั้นที่ 4



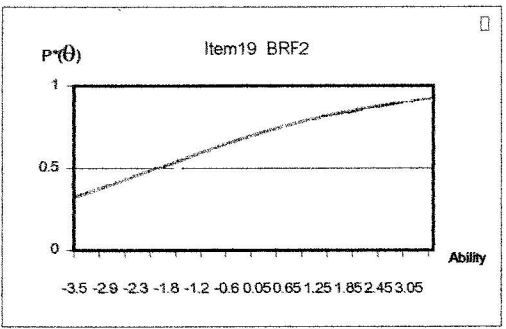
ขอบเขตลำดับขั้นที่ 5



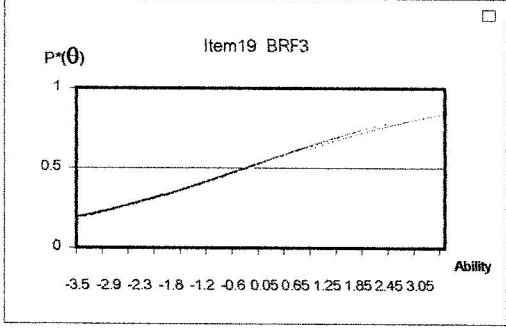
ข้อ 19 ขอบเขตลำดับขั้นที่ 1



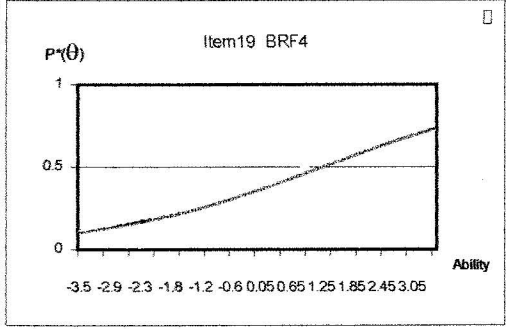
ขอบเขตลำดับขั้นที่ 2



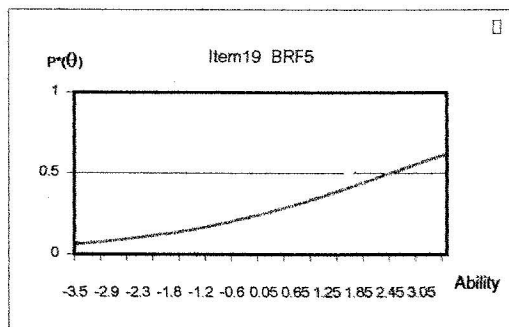
ขอบเขตลำดับขั้นที่ 3



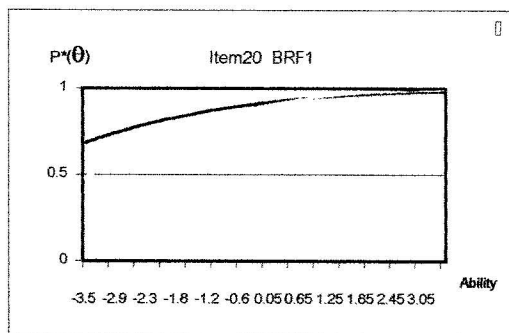
ขอบเขตลำดับขั้นที่ 4



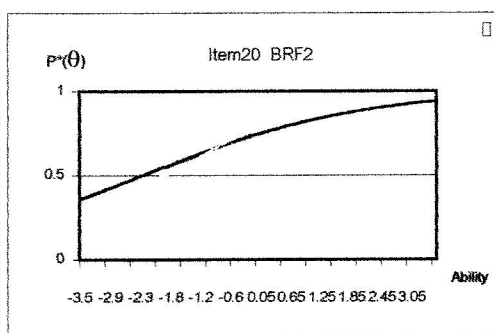
ขอเบตลำนัดบั้นที่ 5



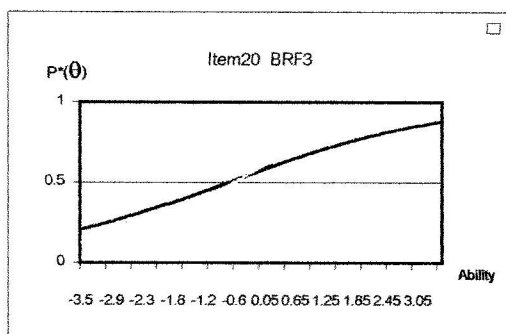
ข้อ 20 ขอเบตลำนัดบั้นที่ 1



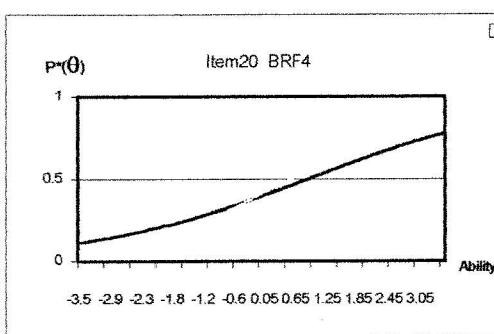
ขอเบตลำนัดบั้นที่ 2



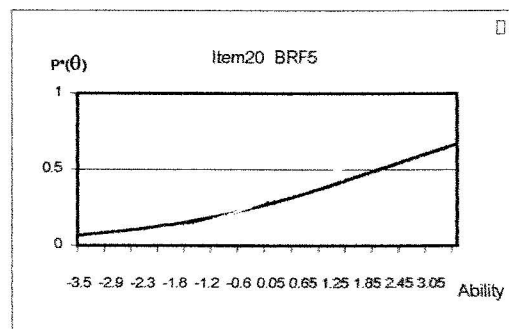
ขอเบตลำนัดบั้นที่ 3



ขอเบตลำนัดบั้นที่ 4



ขอเบตลำนัดบั้นที่ 5



ประวัติย่อผู้วิจัย

ชื่อ-ชื่อสกุล	นางสาวอุทัยวรรณ สายพัฒนา
วัน เดือน ปีเกิด	4 มีนาคม 2498
สถานที่อยู่ปัจจุบัน	802/527 หมู่บ้านริเวอร์ปาร์ควังทอง ตำบลคูคต อำเภอลำลูกกา จังหวัดปทุมธานี 12130
ตำแหน่งหน้าที่การงานในปัจจุบัน	อาจารย์ 2 ระดับ 7
สถานที่ทำงานปัจจุบัน	โรงเรียนธรรมศาสตร์คลองหลวงวิทยาคม อำเภอคลองหลวง จังหวัดปทุมธานี 12120

ประวัติการศึกษา

พ.ศ. 2512	มัธยมศึกษาปีที่ 3 จากโรงเรียนจอมสุรางค์อุปถัมภ์ จังหวัดพระนครศรีอยุธยา
พ.ศ. 2515	ป.กศ. ต้น จากวิทยาลัยครูพระนครศรีอยุธยา
พ.ศ. 2519	กศ.บ. (คณิตศาสตร์) จากมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร
พ.ศ. 2526	น.บ. จากมหาวิทยาลัยรามคำแหง
พ.ศ. 2539	กศ.ม. (การวัดผลการศึกษา) จากมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร
พ.ศ. 2547	กศ.ด. (การทดสอบและวัดผลการศึกษา) จากมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร