



รายงานการวิจัย

4
1754

การวิเคราะห์ความลำเอียงของข้อทดสอบภาษาอังกฤษ
เข้ามหาวิทยาลัย ปี พ.ศ. 2531 - 2533

โดย

รองศาสตราจารย์ ดร. สุทัศน์ ฤกษ์มณี

พ
ภษ 15
008638

สถาบันภาษา
จุฬาลงกรณ์มหาวิทยาลัย

พ.ศ. 2534



งานวิจัยนี้ได้รับเงินอุดหนุนจากงบประมาณแผ่นดิน ปี พ.ศ. 2534
เป็นจำนวนเงิน 30,000 บาท (สามหมื่นบาทถ้วน)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รายงานการวิจัย

เรื่อง



การวิเคราะห์ความลำเอียงของข้อทดสอบภาษาอังกฤษ
เข้ามหาวิทยาลัย ปี พ.ศ. 2531-2533

โดย

รองศาสตราจารย์ ดร. สุวัฒน์ สุขมณีนันต์

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สถานึกษา
จุฬาลงกรณ์มหาวิทยาลัย
พ.ศ. 2534



บทคัดย่อ

ชื่อโครงการ การวิเคราะห์ความลำเอียงของข้อทดสอบภาษาอังกฤษ เข้ามหาวิทยาลัย
ปี พ.ศ. 2531-2533
ชื่อผู้วิจัย รองศาสตราจารย์ ดร. สุวัฒน์ สุขมลสันต์
ปี พ.ศ. 2534
เงินทุน งบประมาณแผ่นดิน ปี พ.ศ. 2534

วัตถุประสงค์หลักของการวิจัยนี้ คือ 1) เพื่อวิเคราะห์หาความลำเอียงต่อเพศ และต่อภาค
ภูมิศาสตร์ของข้อทดสอบภาษาอังกฤษ เข้ามหาวิทยาลัย ชุด กช. และ กชค. ปี พ.ศ. 2531-2533
3 วิธี คือ Delta Plot Method, Chi-square Method และ Three-parameter
Logistic Method และ 2) เพื่อเปรียบเทียบจำนวนข้อทดสอบที่ลำเอียงจากการวิเคราะห์ทั้ง
3 วิธี ประชากรของการวิจัยได้แก่ผู้สอบแบบทดสอบดังกล่าว จำนวน 6 กลุ่ม ๆ ละประมาณ
30,000 - 80,000 คน และแต่ละกลุ่มแบ่งตามเพศ และภาคภูมิศาสตร์ของผู้สอบ ผลวิจัยได้จาก
การสุ่มอย่างง่ายจากประชากรแต่ละกลุ่มโดยกำหนดให้แต่ละกลุ่มมีเพศชายและหญิงเท่ากัน และไม่
เกิน 3,000 คน ทำให้ได้ผลวิจัยกลุ่มละ 424 - 3,000 คน ซึ่งมากกว่าขนาดกลุ่มตัวอย่างที่
เหมาะสม เครื่องมือที่ใช้คือผลการสอบข้อทดสอบภาษาอังกฤษ ชุด กช. และ กชค. จำนวน 600
ข้อ การวิเคราะห์ข้อมูลใช้สถิติหลายอย่าง เช่น Classical Item Analysis, Logistic
Item Analysis, ICC-test, Factor Analysis, Chi-square test, Bias test และ
F-test เป็นต้น

ผลการวิจัยสรุปได้ดังนี้

1. แบบทดสอบภาษาอังกฤษ เข้ามหาวิทยาลัย ชุด กช. ปี 2531-2533 มีความลำเอียง
ต่อเพศ โดยเฉลี่ยประมาณฉบับละ 7-28 ข้อ (ร้อยละ 7-28) ส่วนชุด กชค. มี
ความลำเอียงประมาณฉบับละ 4-41 ข้อ (ร้อยละ 4-41) แล้วแต่วิธีวิเคราะห์
2. แบบทดสอบภาษาอังกฤษ เข้ามหาวิทยาลัย ชุด กช. ปี 2531-2533 มีความลำเอียง
ต่อภาคภูมิศาสตร์โดยเฉลี่ยประมาณฉบับละ 6-45 ข้อ (ร้อยละ 6-45) ส่วนชุด
กชค. มีความลำเอียงประมาณฉบับละ 5-43 ข้อ (ร้อยละ 5-43) แล้วแต่
วิธีวิเคราะห์
3. ข้อทดสอบภาษาอังกฤษ เข้ามหาวิทยาลัย ชุด กช. และ กชค. ปี พ.ศ.
2531-2533 มีความลำเอียงต่อผู้สอบจากภาคอื่นมากกว่าภาคกลาง ประมาณ 2-3
เท่า และมีแนวโน้มว่ามีความลำเอียงต่อผู้สอบเพศชายมากกว่าเพศหญิง

4. การวิเคราะห์หาความลำเอียงของข้อทดสอบด้วย 3 วิธี พบข้อทดสอบที่ลำเอียงต่อเพศและต่อภาคภูมิศาสตร์ของผู้สอบจำนวนแตกต่างกันอย่างมีนัยสำคัญ แต่จำนวนข้อทดสอบที่ลำเอียงของแต่ละวิธีมีความสัมพันธ์กันอย่างไม่มีนัยสำคัญ การวิเคราะห์โดย Three-parameter Logistic Method เมื่อไม่ได้วิเคราะห์ความลำเอียงระดับต่ำพบจำนวนข้อทดสอบที่ลำเอียงมากที่สุด รองลงมาได้แก่ Chi-square Method และ Delta Plot Method พบข้อทดสอบที่ลำเอียงจำนวนน้อยที่สุด

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Project Title: Item Bias Analyses of the English Entrance Examination
in 1988-1990

Researcher: Associate Professor Dr. Suphat Sukamolson

Year: 1991

Grant: Thai Government Budget 1991



The main purposes of this study were 1) to analyze the items biased against sexes and geographical regions of the English Entrance Examinations, Form AB's and ABC's in 1988-1990 by Delta-Plot Method, Chi-square Method and Three-parameter Logistic Method, and 2) to compare the numbers of biased items resulted from the 3 mentioned methods. The populations of this study were 6 groups of approximately 30,000-80,000 actual testees and each was subdivided according to sexes and their geographical regions. They were then sampled by a simple sampling technique with 2 conditions: equal numbers of males and females and not more than 3,000 in each subgroup. The subjects in each subgroup were then ranging from 424 to 3,000 which were more than optimal sizes. The instruments used were 600 items of the 1988-1990 English Entrance Examinations, Form AB's and ABC's. The data were analyzed by means of Classical Item Analysis, Logistic Item Analysis, ICC test, Factor Analysis, Chi-square test, Bias test and F-test.

The findings can be summarized as follows:

1. On average, there were approximately 5-35 items biased against sexes in each form of the examinations (approx. 5%-35%) depending on methods of analyses.
2. On average, there were approximately 8-36 items biased against geographical regions in each form of the examinations (approx. 8%-36%) depending on methods of analyses.

3. Generally speaking, the examinations in 1988-1990 biased against the testees from 4 geographical regions more than those from the central region approximately 3-4 times. There was a tendency that they biased against males more than females.
4. The numbers of biased items resulted from the 3 methods were significantly different and correlated each other insignificantly. Three-parameter Logistic Method when excluded low biases was the most sensitive method. It detected the highest numbers of biased items and the next one was Chi-square Method. Delta-Plot Method was a very conservative method. It detected biased items the least.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



กิตติกรรมประกาศ

ผู้วิจัยใคร่ขอขอบคุณคณะกรรมการวิจัยและคณะกรรมการประจำของสถาบันภาษาที่ได้ช่วยพิจารณาอนุมัติเบื้องต้นให้โครงการวิจัยนี้เป็นโครงการวิจัยของสถาบันภาษา และโดยเฉพาะอย่างยิ่งขอขอบคุณคณะกรรมการวิจัยที่ได้มีส่วนแสดงความคิดเห็นเพิ่มเติม เพื่อช่วยให้งานวิจัยนี้มีความสมบูรณ์และถูกต้องมากยิ่งขึ้น

นอกจากนี้ ผู้วิจัยขอขอบคุณสถาบันบริการคอมพิวเตอร์เป็นอย่างสูง โดยเฉพาะอย่างยิ่งผู้อำนวยการสถาบัน (รองศาสตราจารย์ ดร. ไกรวิชิต ตันติเมท) คุณลีนวน หิรัญวงษ์ คุณเพชรรัตน์ อารีย์ลักษณ์ คุณอภิสิทธิ์ วีระไวทยะ และคุณสมบูรณ์ เกตุพระนิมิตร ที่ให้ความช่วยเหลือและอำนวยความสะดวกต่าง ๆ ในการใช้เครื่องคอมพิวเตอร์เป็นอย่างดีตลอดระยะเวลา 2 เดือนที่ผู้วิจัยทำการวิเคราะห์ข้อมูล ซึ่งมีจำนวนมากที่สุด ยุ่งมากที่สุด ทำให้เสียเวลามากที่สุด และเสียค่าใช้จ่ายมากที่สุด เท่าที่เคยทำการวิจัยมา เนื่องจากต้องวิเคราะห์ข้อมูลจำนวนมากและหลาย ๆ ด้าน โดยเฉพาะการวิเคราะห์ข้อทดสอบด้วย LOGIST ดังนั้นหากไม่ได้รับความช่วยเหลือและความสะดวกในการใช้เครื่องคอมพิวเตอร์แล้ว งานวิจัยนี้จะไม่สามารถกระทำได้เลย

ท้ายสุด ผู้วิจัยขอขอบคุณ คุณจรัสศรี ศรีสัญญาพาณิชย์ เป็นอย่างมากที่ช่วยกรุณาจัดพิมพ์รายงานนี้ให้เสร็จในเวลาที่กำหนดได้อย่างรวดเร็ว และเรียบร้อยดี ผู้วิจัยหวังเป็นอย่างยิ่งว่าผลการวิจัยนี้คงจะเป็นประโยชน์มากสำหรับคณะกรรมการสร้างข้อทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย และผู้ที่เกี่ยวข้อง หรือผู้ที่สนใจในวงการทดสอบทั่วไป ในการที่จะสร้างและปรับปรุงข้อสอบให้มีคุณภาพยิ่งขึ้น

สุทัศน์ สุกมลสันต์

15 กรกฎาคม 2534

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ก
บทคัดย่อภาษาอังกฤษ	ค
กิตติกรรมประกาศ	จ
สารบัญเรื่อง	ฉ
สารบัญตาราง	ช
บทที่ 1: บทนำ	1
- ความเป็นมาและความสำคัญของปัญหา	1
- วัตถุประสงค์ของการวิจัย	2
- สมมติฐานของการวิจัย	3
- ขอบเขตของการวิจัย	4
- ความจำกัดของการวิจัย	4
- คำนิยามศัพท์ที่ใช้ในการวิจัย	6
- ประโยชน์ที่คาดว่าจะได้รับ	8
บทที่ 2: วรรณคดีที่เกี่ยวข้อง	9
- ความลำเอียงของข้อทดสอบ	9
- การวิเคราะห์หาความลำเอียงของข้อทดสอบ 3 วิธี:	11
- Delta Plot Method	11
- Chi-square Method	12
- Three-parameter Logistic Method	13
- งานวิจัยที่เกี่ยวข้องกับการวิเคราะห์หาความลำเอียง	23
- งานวิจัยในต่างประเทศที่ใช้ข้อมูลจริง	23
- งานวิจัยในต่างประเทศที่ใช้ข้อมูลจำลอง	26
- งานวิจัยในประเทศ	28
บทที่ 3: วิธีดำเนินงานวิจัย	31
- ประชากร และผลวิจัย	31
- เครื่องมือ	34

	หน้า
- การเก็บรวบรวมข้อมูล	34
- การวิเคราะห์ข้อมูล	34
- สถิติที่ใช้	39
บทที่ 4: ผลการวิเคราะห์ข้อมูล	40
- ความลำเอียงต่อเพศและต่อภาคภูมิศาสตร์	40
- วิเคราะห์โดย Delta Plot Method	40
- วิเคราะห์โดย Chi-square Method	49
- วิเคราะห์โดย Three-parameter Logistic Method	63
- เปรียบเทียบจำนวนข้อทดสอบที่ล่าเอียงต่อเพศและต่อภาคภูมิศาสตร์	78
- จำนวนข้อที่ล่าเอียงต่อเพศ	78
- จำนวนข้อที่ล่าเอียงต่อภาค	81
บทที่ 5: สรุปผล อภิปรายผล และข้อเสนอแนะ	84
- สรุปผลการวิจัย	88
- อภิปรายผลการวิจัย	89
- ข้อเสนอแนะ	98
บรรณานุกรม	101
ภาคผนวก ก: ผลการวิเคราะห์ด้วยวิธีที่ 1	
ภาคผนวก ข: ผลการวิเคราะห์ด้วยวิธีที่ 2	
ภาคผนวก ค: ผลการวิเคราะห์ด้วยวิธีที่ 3	
ภาคผนวก ง: สถิติ โปรแกรม และข้อมูลบางส่วนที่เกี่ยวข้อง	

เลขหมู่ กพ ๗๕ 15
เลขทะเบียน 008638
วัน, เดือน, ปี 14 ก.พ. 39

สารบัญตาราง

ตารางที่		หน้า
1	การเปรียบเทียบวิธีวิเคราะห์ความลำเอียงของข้อทดสอบ 3 วิธี	20
2	จำนวนประชากรและพลวิสัยของผู้สมัครสอบภาษาอังกฤษเข้า มหาวิทยาลัย ชุด กข. ปี พ.ศ. 2531-2533	32
3	จำนวนประชากรและพลวิสัยของผู้สมัครสอบภาษาอังกฤษเข้า มหาวิทยาลัย ชุด กขค. ปี พ.ศ. 2531-2533	33
4	จำนวนข้อทดสอบที่มีความลำเอียงต่อเพศของแบบทดสอบภาษาอังกฤษ ชุด กข. ปี พ.ศ. 2531-2533 เมื่อวิเคราะห์ด้วย Delta Plot Method	41
5	จำนวนข้อทดสอบที่มีความลำเอียงต่อเพศของแบบทดสอบภาษาอังกฤษ ชุด กขค. ปี พ.ศ. 2531-2533 เมื่อวิเคราะห์ด้วย Delta Plot Method	42
6	จำนวนข้อทดสอบที่มีความลำเอียงต่อผู้สอบจากภาคภูมิศาสตร์ต่าง ๆ ของแบบทดสอบภาษาอังกฤษ ชุด กข. ปี พ.ศ. 2531-2533 เมื่อวิเคราะห์ด้วย Delta Plot Method	43
7	จำนวนข้อทดสอบที่มีความลำเอียงต่อผู้สอบจากภาคภูมิศาสตร์ต่าง ๆ ของแบบทดสอบภาษาอังกฤษ ชุด กขค. ปี พ.ศ. 2531-2533 เมื่อวิเคราะห์ด้วย Delta Plot Method	42
8	จำนวนข้อทดสอบที่มีความลำเอียงต่อเพศของแบบทดสอบภาษาอังกฤษ ชุด กข. ปี พ.ศ. 2531-2533 เมื่อวิเคราะห์ด้วย Chi-square Method	50
9	จำนวนข้อทดสอบที่มีความลำเอียงต่อเพศของแบบทดสอบภาษาอังกฤษ ชุด กขค. ปี พ.ศ. 2531-2533 เมื่อวิเคราะห์ด้วย Chi-square Method	51
10	จำนวนข้อทดสอบที่มีความลำเอียงต่อผู้สอบจากภาคภูมิศาสตร์ต่าง ๆ ของแบบทดสอบภาษาอังกฤษ ชุด กข. ปี พ.ศ. 2531-2533 เมื่อวิเคราะห์ด้วย Chi-square Method	53
11	จำนวนข้อทดสอบที่มีความลำเอียงต่อผู้สอบจากภาคภูมิศาสตร์ต่าง ๆ ของแบบทดสอบภาษาอังกฤษ ชุด กขค. ปี พ.ศ. 2531-2533 เมื่อวิเคราะห์ด้วย Chi-square Method	59

ตารางที่		หน้า
12	จำนวนข้อทดสอบที่มีความลำเอียงต่อเพศของแบบทดสอบภาษาอังกฤษ ชุด กข. ปี พ.ศ. 2531-2533 เมื่อวิเคราะห์ด้วย Three-parameter Logistic Method	64
13	จำนวนข้อทดสอบที่มีความลำเอียงต่อเพศของแบบทดสอบภาษาอังกฤษ ชุด กขค. ปี พ.ศ. 2531-2533 เมื่อวิเคราะห์ด้วย Three-parameter Logistic Method	66
14	จำนวนข้อทดสอบที่มีความลำเอียงต่อผู้สอบจากภาคภูมิศาสตร์ต่าง ๆ ของแบบทดสอบภาษาอังกฤษ ชุด กข. ปี พ.ศ. 2531-2533 เมื่อวิเคราะห์ด้วย (Three-parameter Logistic Model)	68
15	จำนวนข้อทดสอบที่มีความลำเอียงต่อผู้สอบจากภาคภูมิศาสตร์ต่าง ๆ ของแบบทดสอบภาษาอังกฤษ ชุด กขค. ปี พ.ศ. 2531-2533 เมื่อวิเคราะห์ด้วย Three-parameter Logistic Model	73
16	เปรียบเทียบจำนวนข้อทดสอบที่มีความลำเอียงต่อเพศของผู้สอบ เมื่อวิเคราะห์ด้วย 3 วิธี	79
17	เปรียบเทียบจำนวนข้อทดสอบที่มีความลำเอียงต่อภาคภูมิศาสตร์ของผู้สอบเมื่อวิเคราะห์ด้วย 3 วิธี	81
18	จำนวนข้อทดสอบที่มีความลำเอียงต่อเพศ เมื่อวิเคราะห์ด้วย 3 วิธี	90
19	จำนวนข้อทดสอบที่ลำเอียงต่อภาคภูมิศาสตร์เมื่อวิเคราะห์ด้วย 3 วิธี	92

บทที่ 1

บทนำ



ความเข้มงวดและความสำคัญของปัญหา

ในปัจจุบันนี้ ในแต่ละปีการศึกษาจะมีนักเรียนที่เรียนสำเร็จการศึกษาชั้นมัธยมศึกษาตอนปลายแล้ว สมัครเข้าศึกษาต่อในมหาวิทยาลัยของรัฐและเอกชน เป็นจำนวนประมาณ 30,000 คน และจะมีเพียงประมาณร้อยละ 25 เท่านั้น ที่สามารถสอบเข้าศึกษาต่อในมหาวิทยาลัยรับจำกัดของรัฐได้ นอกนั้นก็จะต้องศึกษาในมหาวิทยาลัยรับไม่จำกัด (มหาวิทยาลัยเปิด) และมหาวิทยาลัยหรือวิทยาลัยเอกชน (สถิติการศึกษา, วารสารการศึกษา, 2530: 45) การที่นักเรียนไม่สามารถจะสอบผ่านข้อทดสอบวิชาต่าง ๆ ที่จัดสร้างขึ้นโดยทบวงมหาวิทยาลัยนั้น ส่วนหนึ่งเกิดจากความรู้อันและความสามารถของผู้สมัครสอบมีไม่มากเพียงพอ เมื่อเปรียบเทียบกับผู้ที่สามารถสอบได้ และปัจจัยอีกส่วนหนึ่งที่สำคัญมากก็คือเกิดจากคุณภาพของข้อทดสอบที่นำมาใช้ในการทดสอบ โดยเฉพาะอย่างยิ่งรายวิชาภาษาอังกฤษ เพราะจากการวิเคราะห์ข้อทดสอบเข้ามหาวิทยาลัยวิชาภาษาอังกฤษ ชุด กข. และ กขค. ของผู้รู้ทางการวัดและประเมินผล รวมทั้งทางด้านหลักสูตรในหลายปีที่ผ่านมา พบว่าคุณภาพของข้อทดสอบภาษาอังกฤษ ชุด กข. และ กขค. ในแต่ละปียังมีข้อบกพร่องมาก เช่น ในปี พ.ศ. 2533 Alec Bradford (Bradford, 1933: 8-17) ผู้ซึ่งทำการสอนภาษาอังกฤษ ในมหาวิทยาลัยไทยมานาน พบว่าแบบทดสอบภาษาอังกฤษเข้ามหาวิทยาลัยชุด กข. มีข้อทดสอบที่ไม่มีคำตอบที่ถูกต้องเลย 8 ข้อ มีข้อทดสอบที่มีตัวเลือกที่ถูกต้อง 2 ตัวเลือก จำนวน 11 ข้อ และอีก 1 ข้อ มีตัวเลือกถูกต้องมากถึง 3 ตัวเลือก สรุปแล้ว มีข้อสอบที่มีปัญหาเกี่ยวกับคำตอบที่ถูกต้องรวม 20 ข้อ หรือร้อยละ 20 และยิ่งพบว่าแบบทดสอบภาษาอังกฤษ ชุด กขค. ปี พ.ศ. 2533 มีข้อทดสอบที่มีคำตอบไม่ถูกต้องเลย 8 ข้อ มีข้อทดสอบที่มีคำตอบที่ถูกต้องมากกว่า 1 ตัวเลือก 10 ข้อ ในจำนวนนี้มีอยู่ 3 ข้อที่ตัวเลือกถูกต้องทั้ง 4 ตัวเลือก สรุปแล้วมีข้อทดสอบที่มีปัญหาเกี่ยวกับคำตอบที่ถูกต้องรวม 18 ข้อ หรือร้อยละ 18 จำนวนข้อทดสอบที่มีปัญหาเกี่ยวกับคำตอบที่ถูกต้องที่มีขนาดใกล้เคียงกับผลการศึกษาของสแตนลีย์ สุกมอลล์ ซึ่งทำการศึกษาแบบทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย ปี พ.ศ. 2527 และ 2528 (สแตนลีย์ สุกมอลล์, 2527: 101-114; สแตนลีย์ สุกมอลล์, 2528: 1-20) เช่น พบว่าในปี พ.ศ. 2527 แบบทดสอบภาษาอังกฤษ ชุด กข. และ กขค. มีปัญหาเกี่ยวกับคำตอบที่ถูกต้อง ร้อยละ 15 และ 20 เป็นต้น

ดังนั้น จะเห็นได้ว่าในแต่ละปีแบบทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย ชุด กข. และ กขค. มีข้อผิดพลาดเกี่ยวกับคำตอบที่ถูกต้องฉบับละประมาณร้อยละ 20 ทั้งนี้ยังไม่ได้รวมข้อบกพร่องด้านอื่นที่มีผู้ศึกษาไว้ เช่น เดียวกัน เช่น ความไม่สอดคล้องระหว่างหลักสูตรภาษาอังกฤษระดับมัธยมศึกษา กับ เนื้อหาของแบบทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย เป็นต้น

ข้อบกพร่องต่าง ๆ ของข้อทดสอบดังกล่าวแล้ว ย่อมมีผลโดยตรงต่อผลการสอบเข้ามหาวิทยาลัยของผู้ที่สอบได้และผู้ที่ไม่สอบได้ กล่าวคือทำให้ผลการสอบ และผลการตัดสินการสอบผิดพลาด ผู้ที่สอบได้แท้จริงอาจสอบตก และผู้ที่ไม่สอบได้ก็อาจสอบได้ก็เป็นได้ เพราะคะแนนภาษาอังกฤษ ประมาณ 20 คะแนน มีความสำคัญกับผลการสอบเข้ามหาวิทยาลัยมาก

ดังนั้น จากข้อบกพร่องด้านคุณภาพของข้อทดสอบภาษาอังกฤษ ชุด กข. และ กขค. ดังกล่าวแล้ว ซึ่งอาจถือได้ว่าเป็นเรื่องสำคัญมาก และเป็นปัญหาพื้นฐานที่แก้ได้ง่ายและไม่ควรให้เกิดขึ้น แต่กลับเกิดขึ้นได้ทุกปี ทำให้ผู้วิจัยมีความสงสัยว่าข้อทดสอบดังกล่าวอาจมีความบกพร่องทางด้านอื่นอีก โดยเฉพาะอย่างยิ่งด้านความเป็นอคติหรือความลำเอียงของข้อทดสอบ (item bias) ซึ่งเป็นเรื่องทีละเล็กละน้อย และใหม่กว่า เรื่องคำตอบที่ถูกต้องของข้อทดสอบแต่ละข้อ แต่ก็มีความสำคัญต่อคุณภาพของแบบทดสอบและผลการสอบของผู้สอบแต่ละคน เช่นเดียวกับข้อบกพร่องด้านอื่น ๆ จึงทำการศึกษาเกี่ยวกับ "การวิเคราะห์ความลำเอียงของข้อทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย ปี พ.ศ. 2531-2533" ขึ้น และเนื่องจากแบบทดสอบภาษาอังกฤษชุด กข. และ กขค. เป็นแบบทดสอบที่ใช้กับผู้สมัครสอบทั่วประเทศ โดยมิได้คำนึงถึงความแตกต่างของผู้สมัครสอบทางด้านเพศ วัฒนธรรม สภาพเศรษฐกิจ สังคม ภาษา และเชื้อชาติ ซึ่งล้วนเป็นปัจจัยที่มีผลทำให้ข้อทดสอบเกิดความลำเอียงหรืออคติได้ (Shepard, 1982 : 11; Sinnott, 1980 อ้างจาก Carton and Macro, 1982: 287) แต่ในที่นี้ผู้วิจัยจะขอทำการศึกษาเฉพาะความลำเอียงของข้อสอบต่อความแตกต่างทางเพศ และสถานที่ตั้งทางภูมิศาสตร์ของผู้สมัครสอบเท่านั้น เพราะเป็นปัจจัยที่มีความแตกต่างที่เด่นชัด และง่ายต่อการที่จะนำผลการวิจัยไปใช้ในการปรับปรุงแก้ไขข้อทดสอบที่ลำเอียงให้มีคุณภาพดีขึ้น หากพบว่าข้อทดสอบมีความลำเอียงจริง รวมทั้งการให้ข้อเสนอแนะแก่ผู้ที่เกี่ยวข้องในการสร้างข้อทดสอบภาษาอังกฤษเข้ามหาวิทยาลัยหรือรายวิชาภาษาอังกฤษอื่น ๆ ในการป้องกันความบกพร่องที่ทำให้ข้อทดสอบมีความลำเอียงก็อาจกระทำได้ง่ายในทางปฏิบัติกว่าการแก้ไขความลำเอียงของข้อทดสอบที่เกิดจากปัจจัยอื่น

วัตถุประสงค์ของการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์ดังต่อไปนี้

1. เพื่อวิเคราะห์หาความลำเอียงต่อเพศและต่อภาคภูมิศาสตร์ของข้อทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย ปี พ.ศ. 2531-2533 โดย
 1. วิธีกำหนดจุดค่าเดลต้า (Delta-Plot Method)
 2. วิธีทดสอบความแตกต่างของสัดส่วนระหว่างกลุ่มด้วยไคสแควร์ (Chi-Square Method)

3. วิธีการตอบสนองของข้อทดสอบชนิด 3 พารามิเตอร์ (Three-Parameter Logistic Method)

2. เพื่อเปรียบเทียบจำนวนข้อทดสอบที่มีความลำเอียงต่อเพศและภาคภูมิศาสตร์ เมื่อทำการวิเคราะห์ด้วย 3 วิธีดังกล่าวแล้ว

สมมติฐานของการวิจัย

จากการศึกษาของนักทดสอบและประเมินผลหลายท่าน เช่น Shepard (1982: 11) Simmott (1980 อ้างจาก Carton and Macro, 1982: 237) และ Stricker (1981 อ้างจาก Carton and Macro, 1982: 289) พบว่าแบบทดสอบ GMAT ซึ่งมีบางส่วนเป็น Reading Comprehension นั้น มีความลำเอียงต่อเพศ และจากการศึกษาของรัชชัย เฝ้าพงษ์ (2526 อ้างจาก ทศนีย์ พีรมนตรี, 2530: 12) พบว่าแบบทดสอบวัดความสามารถทางคณิตศาสตร์และภาษาในระดับมัธยมต้นของสำนักงานทดสอบทางการศึกษาและจิตวิทยาของมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร มีความลำเอียงต่อเพศ นอกจากนี้ ทศนีย์ พีรมนตรี (2530: 75) พบว่าแบบทดสอบคณิตศาสตร์ระดับชั้นมัธยมศึกษาปีที่ 6 ของสำนักงานทดสอบทางการศึกษา กรมวิชาการ กระทรวงศึกษาธิการ มีความลำเอียงต่อผู้สอบจากภาคภูมิศาสตร์ต่าง ๆ ในประเทศไทย

นอกจากนี้จากการศึกษาของ ทศนีย์ พีรมนตรี (2530: 75-80) และ สุรศักดิ์ อมรรัตนะศักดิ์ (2532: 80-85) พบว่าวิธีการวิเคราะห์ความลำเอียงของข้อทดสอบต่างวิธีกัน ทำให้ได้ข้อทดสอบที่ลำเอียงจำนวนแตกต่างกัน

ดังนั้น ผู้วิจัยจึงมีสมมติฐานเพื่อการวิจัยครั้งนี้ดังต่อไปนี้

1. แบบทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย ชุด กข. และ กขค. ประจำปี พ.ศ. 2531-2533 น่าจะมีความลำเอียงต่อเพศของผู้สอบ
2. แบบทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย ชุด กข. และ กขค. ประจำปี พ.ศ. 2531-2533 น่าจะมีความลำเอียงต่อภาคภูมิศาสตร์ของผู้สอบ
3. วิธีการวิเคราะห์หาความลำเอียงของข้อทดสอบ 3 วิธีคือ วิธีกำหนดจุดค่าเดลด้า วิธีทดสอบความแตกต่างของสัดส่วนระหว่างกลุ่มด้วยไคสแควร์ และวิธีการตอบสนองของข้อทดสอบชนิด 3 พารามิเตอร์ น่าจะทำให้ได้จำนวนข้อทดสอบที่มีความลำเอียงต่อเพศ และต่อภาคภูมิศาสตร์ของผู้สอบแตกต่างกัน

ขอบเขตของการวิจัย

การวิจัยครั้งนี้มีขอบเขตจำกัดดังต่อไปนี้

1. แบบทดสอบที่นำมาวิเคราะห์ได้แก่ แบบทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย ชุด กข. และ กขค. ประจำปี พ.ศ. 2531-2533 จำนวน 6 ฉบับ และมีข้อทดสอบทั้งหมด 600 ข้อ
2. วิธีการวิเคราะห์ที่หาความลำเอียงของข้อทดสอบใช้เพียง 3 วิธีคือ
 1. วิธีการหาค่าจุดค่าเฉลี่ย
 2. วิธีการทดสอบความแตกต่างของสัดส่วนระหว่างกลุ่มด้วยไคสแควร์
 3. วิธีการทดสอบสองของข้อทดสอบชนิด 3 พารามิเตอร์
3. ผู้สอบแบบทดสอบได้แก่ ผู้สมัครสอบเข้าศึกษาต่อในมหาวิทยาลัย ประจำปีการศึกษา 2531-2533 ทั้งผู้ที่สอบเข้าศึกษาต่อได้ และผู้ที่สอบเข้าศึกษาต่อไม่ได้
4. การวิจัยครั้งนี้ไม่ได้ศึกษาความลำเอียงระดับต่ำ (ลำเอียงน้อย) ของข้อทดสอบเมื่อวิเคราะห์วิธีที่ 3 เพราะคาดว่าจะมีจำนวนมากและ ไม่มีความสำคัญมากนัก

ความจำกัดของการวิจัย

การวิจัยครั้งนี้ แม้ว่าผู้วิจัยจะ ได้พยายามทำให้การวิจัยมีความสมบูรณ์มากที่สุดแล้วก็ตาม แต่ก็มีความจำกัดของการวิจัยบางอย่าง ที่อาจทำให้ผลการวิจัยไม่สมบูรณ์ คือ

1. ผู้วิจัย ไม่อาจทราบได้แน่ชัดว่าแบบทดสอบที่นำมาทำการวิเคราะห์ครั้งนี้ เป็นแบบทดสอบที่มีข้อความตรงกันกับแบบทดสอบต้นฉบับที่ใช้จริงในการทดสอบหรือไม่ เพราะไม่มีแบบทดสอบฉบับที่เป็นทางราชการใช้ ผู้วิจัยได้แบบทดสอบที่นำมาวิเคราะห์ครั้งนี้ จากหนังสือ เฉลยข้อทดสอบเก่าเข้ามหาวิทยาลัยหลายสำนักพิมพ์ และได้ตรวจสอบแล้วว่ามีความตรงกัน
2. ผู้วิจัย ไม่มีคีย์ (key) ตัวเลือกที่ถูกต้องของข้อทดสอบแต่ละข้อที่เป็นเอกสารทางราชการ ผู้วิจัย ได้จัดทำคีย์เฉลยตัวเลือกที่ถูกต้องเอง และได้ตรวจสอบกับผู้รู้ทางภาษาอังกฤษและเจ้าของภาษาหลายท่านรวมทั้งกับคีย์เฉลยที่บางสำนักพิมพ์จัดพิมพ์จำหน่ายด้วย แต่ปรากฏว่าข้อทดสอบบางข้อมีคำตอบที่ถูกต้องมากกว่า 1 ตัวเลือกหรือบางข้อไม่สามารถหาคำตอบที่ถูกต้องได้เลย จึงจำเป็นต้องเลือกตัวเลือกที่ดีที่สุดมาใช้เป็นคีย์เฉลย

3. แนวคิดเชิงทฤษฎีและคำนิยามศัพท์คำว่า "ความลำเอียงของข้อทดสอบ" (item bias) ของการวิเคราะห์หาความลำเอียงของข้อทดสอบแต่ละวิธีแตกต่างกัน และบางวิธี (เช่น วิธีกำหนดจุดค่าเฉลี่ย) ยังไม่มีเกณฑ์มาตรฐานในการตัดสินความลำเอียง ดังนั้น ถ้าวิธีวิเคราะห์ต่างกันหรือเกณฑ์มีการเปลี่ยนแปลงจำนวนข้อทดสอบที่มีความลำเอียงอาจมีการเปลี่ยนแปลงตามไปด้วย
4. การศึกษาความลำเอียงของข้อทดสอบต่อผู้สอบที่อยู่ในภาคภูมิศาสตร์ที่แตกต่างกัน นั้น ผู้วิจัยมุ่งศึกษาระหว่างกลุ่มผู้สอบในภาคกลางกับกลุ่มผู้สอบในภาคภูมิศาสตร์อื่นอีก 4 ภาคเท่านั้น ไม่ได้ทำการศึกษาค้นคว้าความลำเอียงของข้อทดสอบระหว่างกลุ่มผู้สอบในแต่ละภาคภูมิศาสตร์ด้วย
5. การศึกษาครั้งนี้ ไม่ได้มุ่งศึกษาถึงสาเหตุหรือแหล่งที่เกิดของความลำเอียงของข้อทดสอบเป็นรายข้อ เพราะข้อทดสอบมีจำนวนมาก (600 ข้อ) หากแต่จะนำเสนอเป็นบางข้อเพื่อประกอบการอภิปรายผลการวิจัยเท่านั้น
6. เนื่องจากประชากรของการวิจัยครั้งนี้มีขนาดใหญ่มาก กล่าวคือ มีผู้สมัครสอบเข้ามหาวิทยาลัยประมาณปีละ 30,000 คน แต่โปรแกรมคอมพิวเตอร์เพื่อการวิเคราะห์ข้อทดสอบ มีความสามารถจำกัด รวมทั้งการวิเคราะห์ข้อทดสอบรายข้อใช้เวลาในการคำนวณนานและเสียค่าใช้จ่ายสูง โดยเฉพาะเมื่อวิเคราะห์โดยวิธี Logistic Model ผู้วิจัยจึงสุ่มตัวอย่างพลวิจย (subject) มาศึกษากลุ่มละเท่า ๆ กัน และไม่เกิน 1,500 คน เมื่อศึกษาความลำเอียงของข้อทดสอบต่อผู้สอบในภาคภูมิศาสตร์ที่ต่างกัน หรือปีละประมาณ 7,500 คน หรือประมาณร้อยละ 25 ซึ่งเป็นตัวอย่างที่มีขนาดใหญ่อย่างเพียงพอ และเมื่อศึกษาความลำเอียงของข้อทดสอบต่อเพศ ผู้วิจัยสุ่มตัวอย่างพลวิจยมาศึกษากลุ่มละเท่า ๆ กัน และไม่เกิน 3,000 คน หรือปีละประมาณ 6,000 คน หรือประมาณร้อยละ 20 เพราะโปรแกรมวิเคราะห์ข้อทดสอบรายข้อโดยวิธี Classical Model ทำการวิเคราะห์ได้ไม่เกิน 3,000 คนต่อครั้ง

ดังนั้น จะเห็นได้ว่าผู้วิจัยไม่ได้สุ่มตัวอย่างพลวิจยของแต่ละกลุ่ม (เช่น เมื่อแยกเป็นเพศ หรือภาคภูมิศาสตร์) ตามอัตราส่วนของประชากรในแต่ละกลุ่ม แต่ก็เชื่อว่าวิธีการสุ่มอย่างง่ายด้วยเครื่องคอมพิวเตอร์และพลวิจยแต่ละกลุ่มมีขนาดใหญ่เช่นนี้จะทำให้ความผิดพลาดในการวิจัยมีน้อยมาก

คำนิยามศัพท์ที่ใช้ในการวิจัย

แบบทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย ชุด กข. หรือแบบทดสอบภาษาอังกฤษ ชุด กข. หมายถึง แบบทดสอบวิชาภาษาอังกฤษที่สร้างขึ้นโดยคณะกรรมการชุดหนึ่งที่ตั้งขึ้นโดยทบวงมหาวิทยาลัย เพื่อให้ทดสอบผู้สมัครสอบเข้ามหาวิทยาลัยที่ประสงค์จะศึกษาต่อในสายวิทยาศาสตร์และสังคมศาสตร์บางสาขา เนื้อหาของแบบทดสอบครอบคลุมเนื้อหาวิชาภาษาอังกฤษตามหลักสูตรมัธยมศึกษาตอนปลาย และมีกรประกอบด้วย 3 ส่วนคือ English for Every Day Use, English for Future Career และ English for Academic Purposes

แบบทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย ชุด กขค. หรือแบบทดสอบภาษาอังกฤษ ชุด กขค. หมายถึง แบบทดสอบวิชาภาษาอังกฤษที่สร้างขึ้นโดยคณะกรรมการชุดหนึ่งที่ตั้งขึ้นโดยทบวงมหาวิทยาลัย เพื่อให้ทดสอบผู้สมัครสอบเข้ามหาวิทยาลัยที่ประสงค์จะศึกษาต่อในสายมนุษยศาสตร์และสังคมศาสตร์บางสาขา เนื้อหาของแบบทดสอบครอบคลุมเนื้อหาวิชาภาษาอังกฤษตามหลักสูตรมัธยมศึกษาตอนปลาย และมีกรประกอบด้วย 3 ส่วนเช่นเดียวกับแบบทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย ชุด กข. แต่มีอัตราส่วนแตกต่างกัน

ผู้สมัครสอบ หรือผู้สอบหมายถึง นักเรียน นิสิต หรือนักศึกษาที่มีพื้นฐานความรู้อย่างน้อยระดับมัธยมศึกษาตอนปลาย และสมัครสอบวิชาต่าง ๆ เพื่อเข้าศึกษาต่อในมหาวิทยาลัยรับจำกัด (มหาวิทยาลัยปิด) ของรัฐ หรือมหาวิทยาลัยเอกชนบางแห่ง

ความลำเอียงของข้อทดสอบมีความหมายแตกต่างกันดังนี้

1. ความลำเอียงของข้อทดสอบตามการวิเคราะห์ค่าความยาก โดยทฤษฎีประเพณีนิยม (Classical Test Theory) หมายถึงข้อทดสอบที่มีค่าความยาก สำหรับผู้สอบในแต่ละกลุ่มแตกต่างกัน
2. ความลำเอียงของข้อทดสอบตามวิธีกำหนดจุดค่าเดลต้า หมายถึง ข้อทดสอบที่มีค่าเดลต้าของผู้สอบแต่ละกลุ่มเรียงลำดับแตกต่างกัน
3. ความลำเอียงของข้อทดสอบตามวิธีทดสอบความแตกต่างของสัดส่วนระหว่างกลุ่มด้วยไคสแควร์ หมายถึง ข้อทดสอบซึ่งผู้สอบที่ได้คะแนนในช่วงระดับคะแนน

แต่ชวกันมีสัดส่วนของการตอบข้อทดสอบถูกและผิดแตกต่างกัน เนื่องจากมาจากกลุ่มที่แตกต่างกัน

4. ความลำเอียงของข้อทดสอบตามวิธีตอบตนเองของข้อทดสอบชนิด 3 พารามิเตอร์ หมายถึง ข้อทดสอบซึ่งผู้สอบที่มีระดับความสามารถเท่ากันมีโอกาสตอบข้อทดสอบได้ถูกต้องแตกต่างกัน เนื่องจากมาจากกลุ่มที่แตกต่างกัน

เกณฑ์การตัดสินความลำเอียงของข้อทดสอบวิธีต่าง ๆ หมายถึงค่าที่กำกวมขึ้น เพื่อแสดงความลำเอียงของข้อทดสอบตามเกณฑ์ดังนี้

1. ข้อทดสอบใดที่ค่าเอียงและพบตั้งแต่ 1 ภาคขึ้นไป ถือว่าเป็นข้อทดสอบที่มีความลำเอียงต่อภาค และหากว่าพบเฉพาะในกลุ่มเพศใดเพศหนึ่ง ถือว่าเป็นข้อทดสอบที่มีความลำเอียงต่อเพศนั้น
2. ข้อทดสอบใดที่มีระยะตั้งฉากกับค่าคู่อันดับเดลด้า ไปยังเส้นแกนหลัก $> \pm 1.96S_d$ ถือว่าเป็นข้อทดสอบที่มีความลำเอียง ข้อทดสอบที่ค่าระยะห่างมีเครื่องหมาย + หมายความว่าข้อทดสอบนั้นมีความลำเอียงสำหรับกลุ่มผู้สอบที่อยู่ในแกนตั้ง ส่วนข้อทดสอบที่ค่าระยะห่างมีเครื่องหมาย - หมายความว่าข้อทดสอบนั้นมีความลำเอียงสำหรับกลุ่มผู้สอบที่อยู่ในแกนนอน (Angoff, 1982: 107)
3. ข้อทดสอบใดที่ทำให้สัดส่วนการตอบถูกหรือผิดของผู้สอบที่มีระดับคะแนนเดียวกัน แต่มาจากกลุ่มต่างกัน แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ $\alpha = 0.017$ ถือว่าเป็นข้อทดสอบที่มีความลำเอียง (Osterlind, 1983:49) และเมื่อสัดส่วนของผู้ตอบถูกในกลุ่มหนึ่งน้อยกว่าของอีกกลุ่มหนึ่ง แสดงว่าข้อทดสอบข้อนั้นมีความลำเอียงต่อกลุ่มผู้ที่มีสัดส่วนในการตอบถูกน้อยกว่า
4. ข้อทดสอบใดที่มีค่าดัชนีแสดงพื้นที่ระหว่างโค้งลักษณะของข้อทดสอบ (item characteristic curve) มีค่าต่อไปนี้ถือว่าข้อทดสอบมีความลำเอียง (Rudner, 1977 อ้างจาก ทศนิยม นีรมนตรี, 2530: 9) กล่าวคือ
 - 4.1 มีค่าดัชนีแสดงพื้นที่ > 0.70 แสดงว่ามีความลำเอียงมาก
 - 4.2 มีค่าดัชนีแสดงพื้นที่ระหว่าง 0.40 ถึง 0.70 แสดงว่ามีความลำเอียงปานกลาง
 - 4.3 มีค่าดัชนีแสดงพื้นที่ < 0.40 แสดงว่ามีความลำเอียงน้อย

โด่งลักษณะของข้อทดสอบ หมายถึง กราฟที่แสดงความสัมพันธ์ระหว่างระดับความสามารถของผู้สอบกับโอกาสที่จะตอบข้อทดสอบนั้นถูกต้องตามลักษณะเฉพาะของข้อทดสอบนั้น ซึ่งถูกกำหนดด้วยค่าพารามิเตอร์ของข้อทดสอบ

พารามิเตอร์ของข้อทดสอบ หมายถึง ค่าความยาก ค่าอำนาจจำแนก และค่าการเดาของข้อทดสอบแต่ละข้อ ซึ่งคำนวณโดยทฤษฎีการตอบสนองต่อข้อทดสอบ (Item Response Theory)

ภาคภูมิศาสตร์ หรือภาค หมายถึง สถานที่ตั้งของภูมิลำเนาของผู้สมัครสอบที่แสดงไว้ในเอกสารการสมัครสอบคัดเลือกเข้ามหาวิทยาลัย ซึ่งแบ่งออกเป็น 5 ภาค ตามลักษณะที่ตั้งทางภูมิศาสตร์ของทางราชการ คือ ภาคกลาง ภาคตะวันออก ภาคตะวันออกเฉียงเหนือ ภาคเหนือ และภาคใต้

ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อเป็นข้อมูลให้แก่คณะกรรมการหรือผู้ที่รับผิดชอบในการสร้างแบบทดสอบคัดเลือกเข้ามหาวิทยาลัยปีต่อ ๆ ไป ได้ตระหนักถึงลักษณะของข้อทดสอบที่มีอคติ (ความลำเอียง) ในสังคมไทย เพื่อหาทางปรับปรุงแก้ไขข้อทดสอบดังกล่าวให้มีความยุติธรรมแก่ผู้สมัครสอบทั่วไปให้ได้มากที่สุด ซึ่งจะเป็นผลดีต่อผู้สมัครสอบทุกคนและต่อระบบการศึกษาในส่วนรวมได้ด้วย
2. เพื่อเป็นการตรวจสอบคุณภาพของข้อทดสอบภาษาอังกฤษเพื่อคัดเลือกนักเรียนเข้ามหาวิทยาลัยในระยะ 3 ปีที่แล้ว จำนวน 6 ชุด (วิธีหนึ่ง) ว่ามีมากน้อยเพียงใดในด้านความยุติธรรมของข้อทดสอบ
3. เพื่อเป็นการแนะนำวิธีการวิเคราะห์คุณภาพของข้อทดสอบอีกวิธีหนึ่งให้แก่นักการศึกษาทั่วไป ให้เห็นถึงความสำคัญและวิธีการจัดกระทำ เพื่อเป็นแบบอย่างให้แก่ผู้สนใจได้ทำการศึกษาต่อไปสำหรับรายวิชาอื่น ๆ หรือแบบทดสอบอื่น ๆ ได้ด้วย

บทที่ 2

วรรณคดีที่เกี่ยวข้อง

วรรณคดีที่เกี่ยวข้องกับการวิจัยครั้งนี้ อาจจำแนกออกเป็นหัวข้อที่สำคัญ ๆ ได้ดังต่อไปนี้

1. ความลำเอียงของข้อทดสอบ
 2. การวิเคราะห์หาความลำเอียง
 3. งานวิจัยที่เกี่ยวข้องกับการวิเคราะห์หาความลำเอียง
- ต่อไปนี้ คือรายละเอียดของแต่ละหัวข้อยังกล่าวแล้ว

I. ความลำเอียงของข้อทดสอบ

แนวคิดเกี่ยวกับความลำเอียงของข้อทดสอบเป็นแนวคิด ที่นักทดสอบและประเมินผล
มุ่งให้ความสนใจกันอย่างจริงจังเมื่อประมาณปี พ.ศ. 1970 (Berk, 1983: 1) โดยมีความ
เชื่อว่า

1. การศึกษาความลำเอียงของข้อทดสอบ มีประโยชน์ในการพัฒนาแบบทดสอบให้มีคุณภาพ
ดีขึ้นได้ โดยเฉพาะในการลดความลำเอียงของแบบทดสอบให้น้อยลง
2. การศึกษาความลำเอียงของข้อทดสอบสามารถทำไปพร้อม ๆ กับการวิเคราะห์
ข้อทดสอบรายข้อและการสร้างข้อทดสอบในระยะเริ่มแรกอันจะทำให้ข้อทดสอบมี
ความลำเอียงน้อยลงได้
3. การทำให้ข้อทดสอบไม่มีความลำเอียงทำให้แบบทดสอบมีความยุติธรรมสำหรับผู้สอบ
ทั่ว ๆ ไป

อนึ่ง "ความลำเอียง" (bias) กับ "ความไม่ยุติธรรม" (unfairness) เป็นสิ่งที่ไม่
เหมือนกัน กล่าวคือ "ความลำเอียง" เป็นสิ่งที่เกิดขึ้นในข้อทดสอบ หรือแบบทดสอบ แต่ "ความ
ไม่ยุติธรรม" เป็นสิ่งที่เกิดขึ้นในการใช้ข้อทดสอบหรือแบบทดสอบ นักทดสอบทั่วไปให้คำจำกัดความ
คำว่า "ความลำเอียง" คือสิ่งที่ไม่ได้ตั้งใจจะทำการทดสอบอย่างหนึ่ง (a type of invalidity)
ซึ่งเกิดขึ้นในข้อทดสอบหรือแบบทดสอบ (Shepard, 1982: 10) อันมีผลให้ผู้สอบส่วนหนึ่งได้ประโยชน์
จากการสอบ แต่ผู้สอบอีกส่วนหนึ่งเสียผลประโยชน์

ความลำเอียงของข้อทดสอบเกิดจากแหล่งที่สำคัญ 2 แหล่ง คือ

1. เกิดจากการเลือกเนื้อหา (bias in selection) กล่าวคือ
ผู้สร้างข้อทดสอบเลือกเนื้อหาเฉพาะส่วนใดส่วนหนึ่งมาสร้างข้อทดสอบ ทำให้
ข้อทดสอบไม่มีสัดส่วนของเนื้อหาที่สมดุลกัน

2. เกิดจากการสร้างข้อทดสอบ (bias in construction) กล่าวคือผู้สร้างข้อทดสอบใช้ข้อความบางอย่างในข้อทดสอบที่เอื้อให้เกิดประโยชน์แก่ผู้สอบบางส่วน เช่น ใช้ชื่อเครื่องสำอางที่ผู้หญิงชอบใช้ในข้อทดสอบ ทำให้ผู้ชายที่สอบเสียเปรียบ เป็นต้น

แต่อย่างไรก็ตาม ในทางปฏิบัตินั้น ความลำเอียงของข้อทดสอบมีนิยามเชิงปฏิบัติการแตกต่างกันแล้วแต่วิธีการค้นหาความลำเอียง เช่น ในกรณีที่ใช้วิธีทดสอบความแตกต่างของสัดส่วนของแต่ละกลุ่มด้วยไคสแควร์ (Chi-square Method) ก็เชื่อว่า

"ข้อทดสอบที่มีความลำเอียงคือข้อทดสอบซึ่งผู้สอบที่ได้คะแนนในช่วงระดับคะแนนเดียวกัน มีสัดส่วนของการตอบข้อทดสอบนั้นถูกหรือผิดแตกต่างกัน เนื่องจากมาจากกลุ่มที่ต่างกัน" (แปลงจากคำนิยามของข้อทดสอบที่ไม่มีความลำเอียง (unbiased item) ของ Scheunerman, 1975: 2 อ้างจาก Shepard, 1982: 23)

และในกรณีที่ใช้วิธีการตอบสนองของข้อทดสอบชนิด 3 พารามิเตอร์ (Three-Parameter Logistic Method) ก็เชื่อว่า

"ข้อทดสอบที่มีความลำเอียง คือข้อทดสอบซึ่งผู้สอบที่มีระดับความสามารถเท่ากันมีโอกาสตอบข้อทดสอบได้ถูกต้องแตกต่างกัน เนื่องจากมาจากกลุ่มที่ต่างกัน" (แปลงมาจากคำนิยามของข้อทดสอบที่ไม่มีความลำเอียงของ Pine, 1977: 38 อ้างจาก Osterlind, 1983: 55)

นอกจากนี้ เนื่องจากความลำเอียงของข้อทดสอบ เป็นสิ่งที่เกิดขึ้นในข้อทดสอบ การวิเคราะห์หาความลำเอียงที่ดีควรใช้ทั้งการวิเคราะห์เชิงตรรก (Logical Analysis) เพื่อศึกษาความสอดคล้องของสิ่งที่มุ่งทดสอบกับเนื้อหาของข้อทดสอบโดยเฉพาะอย่างยิ่งในระยะแรกของการสร้างข้อทดสอบ และใช้การวิเคราะห์เชิงสถิติ (Statistical Analysis) ด้วย เพื่อตรวจสอบความเป็นเอกพันธ์ของเนื้อหาของข้อทดสอบ (homogeneous sets of items) หลังจากนั้นแล้วก็ควรใช้การวิเคราะห์เชิงตรรกอีกครั้งหนึ่ง เพื่อตรวจสอบกระแสน (pattern) ของการตอบข้อทดสอบแต่ละข้อจากค่าสถิติต่าง ๆ เพื่อแก้ไขให้ข้อทดสอบมีคุณภาพดียิ่งขึ้น

อนึ่ง เนื่องจากการวิเคราะห์หาความลำเอียงของข้อทดสอบ ต้องอาศัยผลการสอบจากข้อสอบข้ออื่น ๆ เพื่อเป็นฐานกำหนดระดับความสามารถของผู้สอบ ดังนั้นการวิเคราะห์ดังกล่าวจะวิเคราะห์ข้อทดสอบเพียงข้อเดียวไม่ได้

II. การวิเคราะห์หาความลำเอียงของข้อทดสอบ 3 วิธี

การวิเคราะห์หาความลำเอียง 3 วิธีที่ใช้ในงานวิจัยนี้ ได้แก่

1. วิธีกำหนดจุดค่าเดลต้า (Delta-Plot Method)
2. วิธีทดสอบความแตกต่างของสัดส่วนระหว่างกลุ่มด้วยไคสแควร์ (Chi-Square Method)
3. วิธีการตอบสนองของข้อทดสอบชนิด 3 พารามิเตอร์ (Three-Parameter Logistic Method)

การวิเคราะห์ทั้ง 3 วิธีดังกล่าวแล้วมีรายละเอียดที่สำคัญดังต่อไปนี้

1. Delta-Plot Method หรือ Transformed Item Difficulty (TID)

วิธีการและแนวคิด

การหาความลำเอียงของข้อทดสอบตามวิธีการนี้จะต้องทำการคำนวณหาค่าความยากง่าย (p) ของข้อทดสอบแต่ละข้อจากผู้สอบแต่ละกลุ่ม ตามวิธีประเพณีนิยม (Classical Model) ค่าความยากง่ายของข้อทดสอบแต่ละข้อจะมี 2 ค่าตามจำนวนกลุ่มที่แบ่ง ในการเปรียบเทียบค่าความยากง่ายนี้จะใช้วิธีการแปลงค่าความยากง่าย (p-values) ไปเป็นค่าซี (z-values) แล้วทำการแปลงค่าซีเป็นค่าเดลต้า (Δ -values) อีกต่อหนึ่งโดยการที่ใช้สมการ $\Delta = 4z + 13$ แล้วนำค่าเดลต้าแต่ละค้อมา Plot ลงบน Bivariate Graph โดยให้แนวนอนแทนค่าเดลต้า สำหรับกลุ่มหนึ่ง และให้แนวตั้งแทนค่าเดลต้าของอีกกลุ่มหนึ่ง จุดต่าง ๆ ที่ปรากฏนี้จะอยู่ในรูปลักษณะที่เป็น Ellipse แล้วทำการคำนวณหาระยะทางที่จุดต่าง ๆ เหล่านี้ห่างจากเส้นแกนหลัก ข้อสอบข้อใดที่ห่างจากเส้นแกนหลัก เกินกว่าเกณฑ์ที่กำหนดไว้ถือว่าเป็นข้อสอบที่มีความลำเอียง (Angoff, 1982: 107)

การทดสอบสมมติฐาน

$$\begin{aligned} H_0 &: \Delta_{i1} - \Delta_{i2} = 0 \\ H_1 &: \Delta_{i1} - \Delta_{i2} \neq 0 \end{aligned}$$

ในเมื่อ Δ_{i1} = ค่าความยากง่ายของข้อทดสอบข้อที่ i ของกลุ่มที่ 1
 Δ_{i2} = ค่าความยากง่ายของข้อทดสอบข้อที่ i ของกลุ่มที่ 2

จากสมมติฐานหมายความว่าข้อทดสอบที่ค่าเฉลี่ยที่มาจากกลุ่มต่างหากว่าไม่เรียงลำดับเหมือนกันแล้ว จะถือว่าเป็นอันตรกิริยา (interaction) เกิดขึ้นระหว่างข้อทดสอบกับกลุ่มที่ต่างกัน ซึ่งถือว่าเป็นข้อทดสอบที่มีความลำเอียง และการจะทดสอบดังกล่าวจะต้องหาระยะห่างตั้งฉากของจุดที่เกิดจากคู่ของค่าเฉลี่ยไปยังแกนหลัก (d_1) แล้วเทียบกับค่าความเบี่ยงเบนมาตรฐาน (S_d) ซึ่งวิธีการภาคคำนวณได้กล่าวไว้ในตอนที่ 3)

2. Chi-square Method หรือ χ^2 - test

วิธีการและแนวคิด

การหาความลำเอียงของข้อทดสอบวิธีนี้ จะต้องแบ่งผู้สอบในแต่ละกลุ่มออกเป็นกลุ่มย่อยโดยอาศัยคะแนนรวมของแบบทดสอบเป็นเกณฑ์ตามระดับความสามารถ แต่ไม่มีเกณฑ์ที่แน่นอนว่าควรแบ่งผู้สอบออกเป็นกี่กลุ่ม เพียงแต่มีข้อเสนอแนะว่าในแต่ละกลุ่มย่อยนี้ ควรมีความถี่ของผู้ตอบถูกหรือผิดไม่น้อยกว่า 5 คนเป็นอย่างน้อย วิธีการหาความลำเอียงโดยใช้ χ^2 - test นี้ มีหลายวิธีแต่ที่นิยมใช้กันมากมี 2 วิธีคือ

1. Marascuilo Method One วิธีนี้จะแบ่งกลุ่มผู้สอบออกเป็นกลุ่มย่อยขนาด 2×2 contingency table แล้วทดสอบสัดส่วนของการตอบถูกหรือผิดของผู้สอบด้วย χ^2 - test หลายครั้งตามจำนวนกลุ่มย่อย แล้วจึงทดสอบความคลาดเคลื่อนชนิดที่ 1 (Type I error) ในภายหลัง รวมทั้งหมดอีกครั้งหนึ่งเพื่อตรวจสอบว่าความลำเอียงของข้อทดสอบเกิดแก่ผู้สอบกลุ่มใดและอยู่ในระดับความสามารถใด

2. Marascuilo Method Three วิธีนี้แม้ว่าจะแบ่งผู้สอบออกเป็นกลุ่มย่อยเช่นเดียวกับวิธีที่ 1 แต่ว่าจะทำการทดสอบสัดส่วนของการตอบถูกหรือผิดของผู้สอบด้วย χ^2 - test เพียงครั้งเดียวเท่านั้น เพื่อตรวจสอบว่าความลำเอียงของข้อทดสอบเกิดแก่ผู้สอบกลุ่มใด

สำหรับการตัดสินใจว่าข้อทดสอบใดมีความลำเอียงจะอาศัยความมีนัยสำคัญทางสถิติของการทดสอบด้วย χ^2 - test เป็นเกณฑ์

การทดสอบสมมติฐาน

วิธีที่ 1:

$$H_0 : P_{1j1} - P_{2j1} = 0$$

$$H_1 : P_{1j1} - P_{2j1} \neq 0$$

ในเมื่อ P_j = สัดส่วนของผู้ตอบข้อทดสอบ ได้ถูกต้องที่ระดับความสามารถ
หนึ่งและจากกลุ่มหนึ่ง

วิธีที่ 2:

$$H_0 : P_{1j1} = P_{2j1} = \dots = P_{1j1}$$

$$H_1 : P_{1j1} \neq P_{2j1} \neq \dots \neq P_{1j1}$$

จากสมมติฐานของวิธีที่ 1 หรือ 2 หมายความว่าข้อทดสอบที่ผู้สอบมาจาก
กลุ่มที่แตกต่างกันแต่มีความสามารถอยู่ในระดับเดียวกัน หากตอบข้อทดสอบนั้นถูกหรือผิด
ในอัตราส่วนที่ต่างกันแล้ว ข้อทดสอบนั้นถือว่ามีความลำเอียง

สำหรับการวิจัยครั้งนี้ ผู้วิจัยเลือกใช้วิธีที่ 1 ซึ่ง วิธีการคำนวณได้กล่าว
ไว้ในบทที่ 3

3 Three-parameter Logistic Method (ICC-3)

การวิเคราะห์หาความลำเอียงวิธีนี้อาศัยแนวคิดจากทฤษฎีการวัดผลแนวใหม่ คือ
ทฤษฎีการตอบสนองต่อข้อทดสอบ (Item Response Theory) ซึ่งมีรายละเอียดพอสังเขปดังนี้
ความเป็นมาของทฤษฎีการตอบสนองต่อข้อทดสอบ

Ferguson และ Lawley เป็นผู้ริเริ่มทฤษฎีการตอบสนองต่อข้อสอบ โดยมีหลักการว่า
ผลการสอบของผู้สอบจากแบบสอบใด ๆ ขึ้นอยู่กับความสามารถ (Ability of Skill) ของผู้สอบ
กับลักษณะเฉพาะของข้อทดสอบ ต่อมาในปี พ.ศ. 1952 Lord ได้เสนอทฤษฎีใหม่ในรูปแบบโค้ง
ลักษณะข้อสอบ (Item Characteristic Curve: ICC) โดย Lord ได้เสนอว่าโค้งลักษณะ
ข้อสอบมีลักษณะเป็นโค้งปกติสะสม ต่อมาเรียกว่า Normal Ogive Model ซึ่งโมเดลนี้จะกล่าว

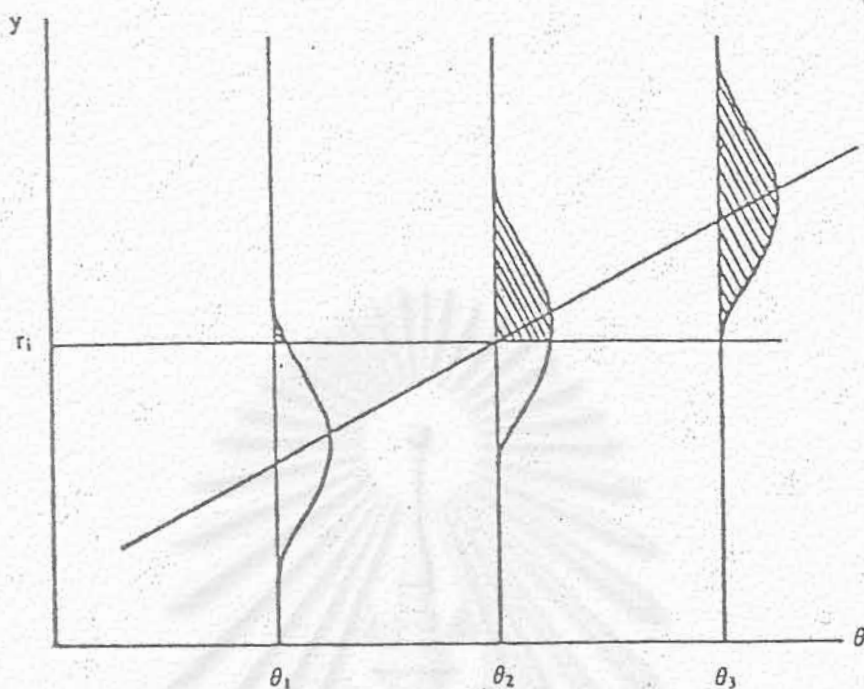
ถึงพารามิเตอร์ 2 ตัว คือ ค่าความยากง่ายและค่าอำนาจจำแนก แต่เนื่องจากโมเดลนี้มีการคำนวณยุ่งยากมาก และขาดแคลนโปรแกรมคอมพิวเตอร์ที่จำเป็นต้องใช้วิเคราะห์ข้อมูลตามทฤษฎี จึงทำให้ Lord หยุดความสนใจในทฤษฎีนี้ไประยะหนึ่ง

ต่อมาในปี ค.ศ. 1960 Rasch ได้ศึกษาเกี่ยวกับเรื่องนี้ โดยไม่ทราบแนวคิดของ Lord มาก่อน โดย Rasch ได้เสนอแนวคิดในรูปของพารามิเตอร์ตัวเดียว คือ ค่าความยากง่าย บางครั้งจึงมีผู้เรียกแบบจำลองนี้ว่า Rasch Model และในปี ค.ศ. 1965 Lord ก็หันมาสนใจและพัฒนาทฤษฎีการตอบสนองต่อข้อทดสอบใหม่ (Warm, 1979: 19)

ต่อมาในปี ค.ศ. 1968 Birnbaum ได้เสนอ Logistic Model ที่ใช้พารามิเตอร์ 2 ตัวคือ ค่าความยากง่ายและค่าอำนาจจำแนก ซึ่งเป็นแบบจำลองที่ง่ายกว่าของ Lord จึงทำให้ Logistic Model เป็นที่นิยมแพร่หลายและมีการพัฒนาขึ้นเรื่อย ๆ จนกระทั่งสามารถใช้ได้กับพารามิเตอร์ตัวเดียว และพารามิเตอร์สามตัว (Warm, 1979: 19-21)

หลักการของทฤษฎีการตอบสนองต่อข้อทดสอบ

ทฤษฎีการตอบสนองต่อข้อทดสอบเป็นทฤษฎีที่อธิบายถึงความสัมพันธ์ระหว่างลักษณะ หรือความสามารถที่มีอยู่ในตัวบุคคลกับพฤติกรรมการตอบสนองต่อข้อทดสอบของบุคคลนั้น โดยทฤษฎีนี้มีความเชื่อว่าพฤติกรรมการตอบสนองต่อข้อทดสอบของบุคคล จะถูกกำหนดโดยลักษณะหรือความสามารถที่มีอยู่ในตัวบุคคล (Lord and Novick, 1968: 358) ซึ่งไม่สามารถจะสังเกตได้ ทฤษฎีนี้พยายามที่จะอธิบายความสัมพันธ์ระหว่างลักษณะภายในตัวบุคคลกับพฤติกรรมที่บุคคลตอบสนองต่อข้อทดสอบ การอธิบายความสัมพันธ์ดังกล่าวแสดงออกมาในรูปแบบจำลองทางคณิตศาสตร์ โดยให้คะแนนที่ได้รับจากการตอบข้อทดสอบ (y) แทนพฤติกรรมการตอบสนองต่อข้อทดสอบ ให้ θ แทนลักษณะหรือความสามารถในตัวบุคคล และ r_i เป็นเกณฑ์ที่บอกว่า y แค่ไหนจึงจะทำข้อทดสอบข้อ i ได้ถูก ดังนั้น ถ้า $y > r_i$ แสดงว่าทำข้อทดสอบข้อ i ได้ถูก และถ้า $y < r_i$ แสดงว่าทำข้อทดสอบข้อ i ผิด ความสัมพันธ์ระหว่างความสามารถของผู้สอบ (θ) กับพฤติกรรมในการตอบสนองต่อข้อทดสอบ (y) แสดงได้ดังแผนภูมิตี่ 1



แผนภูมิที่ 1: ความสัมพันธ์ระหว่างความสามารถของผู้สอบกับ
พฤติกรรมกรรมการตอบสนองต่อข้อทดสอบ

จากแผนภูมิที่ 1 จะเห็นได้ว่า ถ้านำโอกาสที่จะตอบถูก (เงินที่ส่วนที่รเงา) ในระดับความสามารถต่าง ๆ มาเขียนกราฟใหม่ จะได้โค้งลักษณะของข้อทดสอบ (ICC) เป็นรูปต่าง ๆ ซึ่งขึ้นอยู่กับฟังก์ชันทางคณิตศาสตร์และจำนวนพารามิเตอร์ที่จะใช้อธิบาย ฟังก์ชันทางคณิตศาสตร์ที่แสดงความสัมพันธ์ระหว่างระดับความสามารถกับพฤติกรรมกรรมการตอบสนองต่อข้อทดสอบเรียกว่า ฟังก์ชันการตอบสนองต่อข้อทดสอบ (Item Characteristic Function) (Lord and Novick, 1968: 360) ซึ่งสามารถเขียนฟังก์ชันได้ดังนี้

$$P_1(\theta) = \text{Prob}(U_1 = 1 | \theta) \text{ เมื่อ } U_1 = 0, 1$$

จากฟังก์ชันข้างต้นนี้ หมายถึงโอกาสที่ผู้สอบซึ่งมีความสามารถ θ จะตอบคำถามข้อ 1 ได้ถูกต้องหรือไม่

ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองต่อข้อทดสอบ

ทฤษฎีการตอบสนองต่อข้อทดสอบนั้น มีข้อตกลงเบื้องต้นดังนี้

1. ความมีมิติเดียว (Unidimensionality) ของแบบทดสอบ ซึ่งเชื่อว่าข้อทดสอบที่ดี ในแบบทดสอบควรมีลักษณะเป็นเอกพันธ์ นั่นคือแบบสอบนั้นจะต้องมุ่งวัดความสามารถอย่างใดอย่างหนึ่งเพียงความสามารถเดียว หากไม่กำหนดข้อตกลงเบื้องต้นเช่นนี้ จะทำให้แบบจำลองของทฤษฎีมีความยุ่งยากมาก (Hambleton and Cook, 1977: 77) ส่วนการตรวจสอบว่าข้อมูลจากการสอบเป็นไปตามข้อตกลงเบื้องต้นหรือไม่ อาจทำได้โดยการวิเคราะห์องค์ประกอบ (Factor Analysis) และวิธีอื่น ๆ เช่น การหาค่าสัมประสิทธิ์สหสัมพันธ์อย่างง่ายแบบเพียร์สัน (Pearson Product-moment Correlation) ระหว่างค่าไบซีเรียล (biserial correlation) กับค่าน้ำหนักของตัวประกอบที่ 1 ของข้อทดสอบแต่ละข้อ เป็นต้น (Warm, 1978: 104)
2. ความเป็นอิสระของข้อทดสอบ (Local Independence) กล่าวคือการตอบสนองต่อข้อทดสอบข้อใดข้อหนึ่งของข้อสอบ ไม่มีผลต่อการตอบสนองต่อข้อทดสอบข้ออื่น ๆ ในแบบทดสอบ ไม่ว่าข้อทดสอบเหล่านั้นจะอยู่ในส่วนใดของแบบทดสอบ ในการตรวจสอบว่าข้อทดสอบแต่ละข้อเป็นไปตามข้อตกลงนี้หรือไม่ ทำได้โดยอาศัยเทคนิคการวิเคราะห์องค์ประกอบเช่นเดียวกัน (Hambleton and Cook, 1977: 78)
3. ข้อทดสอบแต่ละข้อมีโค้งลักษณะของข้อทดสอบ (Item Characteristic Curve) กล่าวคือ โค้งลักษณะข้อสอบเป็นกราฟของฟังก์ชันทางคณิตศาสตร์ที่แสดงถึงความสัมพันธ์ระหว่างโอกาสในการตอบข้อทดสอบนั้นถูกต้องกับความสามารถหรือลักษณะที่วัดโดยข้อสอบข้อนั้น โค้งลักษณะของข้อทดสอบมีหลายรูปแบบ ทั้งขึ้นอยู่กับแบบจำลองที่ใช้อธิบายความสัมพันธ์ดังกล่าว

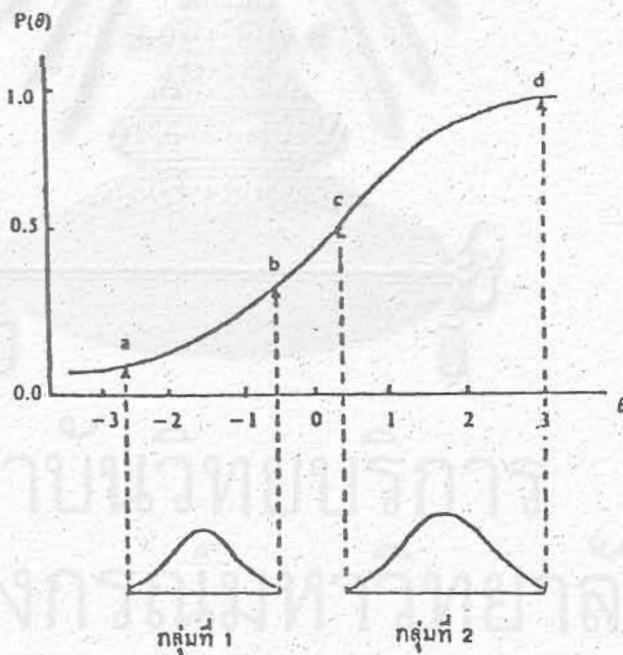
ลักษณะเด่นของทฤษฎีการตอบสนองต่อข้อทดสอบ

ทฤษฎีการตอบสนองต่อข้อทดสอบมีความดีเหนือกว่า (Superiority) ทฤษฎีสอบแบบประเพณีนิยม (Classical Test Theory) หลายอย่าง เช่น

1. ความไม่แปรเปลี่ยนของค่าพารามิเตอร์ (Invariant of Item Parameter) กล่าวคือ ไม่ว่าจะประมาณค่าจากกลุ่มตัวอย่างที่มีความสามารถระดับใดก็ตาม ค่าพารามิเตอร์จะไม่แปรเปลี่ยนไปตามกลุ่มของผู้สอบ นั่นคือ

- ค่าความยากง่าย (b) ค่าอำนาจจำแนก (a) และค่าโอกาสในการตอบถูก โดยการเดา (c) จะไม่แปรเปลี่ยนไม่ว่าจะประมาณค่าจากผู้สอบกลุ่มใด
2. สามารถใช้ข้อทดสอบกับใครก็ได้ (Person-free) นั่นคือ ไม่ว่าจะนำข้อทดสอบไปใช้สอบกับบุคคลกลุ่มใด ได้ังลักษณะข้อทดสอบก็จะคงเดิม
 3. ในการประมาณค่าความสามารถแท้ (θ) ของผู้สอบ สามารถใช้ข้อทดสอบข้อใดก็ได้ (Item-free) จำนวนเท่าใดก็ได้ บางครั้งอาจใช้ข้อทดสอบเพียง 3-5 ข้อ ก็สามารถประมาณค่าความสามารถแท้ (true ability) ของผู้สอบได้แล้ว

ความไม่แปรเปลี่ยนของค่าพารามิเตอร์นั้นสามารถอธิบายได้ด้วยแผนภูมิที่ 2 ต่อไปนี้



แผนภูมิที่ 2: ความไม่แปรเปลี่ยนของค่าพารามิเตอร์ของข้อทดสอบที่ประมาณค่าจากกลุ่มที่แตกต่างกัน

จากแผนภูมิที่ 2 แสดงให้เห็นว่าในการประมาณค่าจากกลุ่มผู้สอบที่มีความสามารถต่ำ (กลุ่มที่ 1) เส้นโค้งลักษณะของข้อทดสอบที่ได้จะอยู่ในช่วง a-b เส้นโค้งลักษณะข้อทดสอบในช่วงนี้จะอธิบายถึงโอกาสในการตอบข้อทดสอบข้อนี้ถูกของผู้สอบที่มีความสามารถต่ำ และทำนองเดียวกัน เส้นโค้งลักษณะของข้อทดสอบในช่วง c-d ก็แสดงถึงโอกาสที่จะตอบข้อทดสอบข้อนี้ถูกของผู้สอบที่มีความสามารถสูง (กลุ่มที่ 2) ถ้าข้อทดสอบข้อนี้วัดความสามารถอันเดียวกันของกลุ่มตัวอย่างทั้งสองกลุ่มแล้ว เส้นโค้งลักษณะของข้อสอบย่อมมีเส้นเดียว เส้นโค้งในช่วง a-b และช่วง c-d ย่อมเป็นส่วนหนึ่งของเส้นโค้งลักษณะข้อทดสอบข้อนี้ เมื่อเส้นโค้งลักษณะของข้อสอบมีได้เส้นเดียว ตัวพารามิเตอร์ซึ่งเป็นตัวกำหนดรูปร่างลักษณะของเส้นโค้งก็ย่อมเป็นค่าเดียวกัน

Lord (1980: 36) ได้อธิบายให้เห็นว่าการไม่แปรเปลี่ยนของค่าพารามิเตอร์นั้น ไม่ได้หมายความว่าค่าพารามิเตอร์ที่ได้จากการประมาณค่าโดยใช้กลุ่มตัวอย่างที่แตกต่างกันจะมีค่าเท่ากันเสมอ ค่าพารามิเตอร์ของข้อทดสอบจะเท่ากันหรือไม่ ขึ้นอยู่กับเงื่อนไขในการทดสอบบางประการ เช่น ถ้าเลือกมาตรวัดที่มีจุดเริ่มต้นเดียวกันและหน่วยในการวัดหน่วยเดียวกันแล้วค่าพารามิเตอร์ที่ประมาณค่าจากกลุ่มตัวอย่างที่แตกต่างกันจะมีค่าเท่ากัน ในทำนองเดียวกันหากเลือกมาตรวัดที่มีจุดเริ่มต้นและมีหน่วยในการวัดแตกต่างกันแล้ว การไม่แปรเปลี่ยนของค่าพารามิเตอร์หมายความว่าค่าพารามิเตอร์ที่ประมาณได้จากกลุ่มตัวอย่างที่ต่างกันของข้อทดสอบชุดหนึ่งจะมีความสัมพันธ์กันในเชิงเส้นตรง ซึ่งเมื่อปรับเปลี่ยน (transform) แล้วจะมีค่าเท่าเดิมหรือใกล้เคียงกันมาก

นอกจากนี้ Baker (1977: 170) ยังได้อธิบายเพิ่มเติมเกี่ยวกับการไม่แปรเปลี่ยนของค่าพารามิเตอร์นี้ว่าขึ้นอยู่กับเงื่อนไข 2 ประการคือ ความสามารถของผู้สอบ (θ) ที่กล่าวถึงต้องสามารถนิยามได้ชัดเจนและวัดได้ด้วยข้อสอบ และความสามารถที่วัดนั้นจะต้องมีความคงที่ภายในช่วงระยะเวลาหนึ่งด้วย

วิธีการและแนวคิดเกี่ยวกับทฤษฎีการวัดความถนัดเชิง

จากแนวคิดดังกล่าวมาแล้ว จะเห็นได้ว่าทฤษฎีการตอบสนองต่อข้อสอบเป็นทฤษฎีทางการทดสอบที่มีข้อตกลงเบื้องต้นที่ค่อนข้างจะเข้มงวดกว่าทฤษฎีการวัดแบบประเพณีนิยม ทฤษฎีนี้มุ่งที่จะอธิบายหรือทำนายความสามารถของผู้สอบซึ่งเป็นคุณลักษณะที่ไม่สามารถวัดได้โดยตรง โดยอาศัยผลการสอบหรือคะแนนที่ผู้สอบแสดงออกมาให้เห็น ในการอธิบายความสัมพันธ์ระหว่างโอกาสที่จะตอบข้อสอบถูกกับระดับความสามารถนี้จะใช้โค้งลักษณะของข้อทดสอบ (ICC) และลักษณะของโค้งนี้ จะถูกกำหนดโดยค่าพารามิเตอร์ที่เกี่ยวกับความยากง่าย (b) ค่าอำนาจจำแนก (a) และค่าการเดา (c) โค้งลักษณะของข้อทดสอบนี้เป็นดัชนีชี้ถึงโอกาสที่ผู้สอบที่มีความสามารถระดับหนึ่งจะตอบข้อสอบได้ถูกมากขึ้นเพียงใดและโอกาสที่ตอบข้อสอบถูกนั้นจะเป็นเพิ่มขึ้นตามลำดับความสามารถที่เพิ่มขึ้น

ดังนั้นวิธีการวิเคราะห์หาความลำเอียงโดยวิธี Three-Parameter Logistic Method ถือว่าข้อทดสอบมีความลำเอียงก็ต่อเมื่อผู้สอบที่มีระดับความสามารถเท่ากันแต่อยู่ต่างกลุ่มกัน มีโอกาสจะตอบถูกไม่เท่ากัน ซึ่งความลำเอียงของข้อทดสอบนี้อาจแสดงให้เห็นได้ด้วย โคน์ลักษณะของข้อทดสอบ กล่าวคือ ถ้าผู้สอบต่างกลุ่มกันมีโคน์ลักษณะของข้อทดสอบข้อเดียวกันแตกต่างกัน ข้อทดสอบข้อนั้นถือได้ว่าเป็นข้อทดสอบที่มีความลำเอียง ในทำนองเดียวกัน หากโคน์ลักษณะของข้อทดสอบของผู้สอบต่างกลุ่มกันมีลักษณะ โคน์เหมือนกันก็ย่อมแสดงว่าข้อทดสอบข้อนั้นไม่ลำเอียง ส่วนวิธีการในการประมาณค่าพารามิเตอร์ 3 ตัวนั้น มักนิยมใช้เครื่องคอมพิวเตอร์โดยอาศัยโปรแกรมชื่อ Logist

การทดสอบสมมติฐาน

การทดสอบสมมติฐานทางสถิติของการวิเคราะห์ความลำเอียง โดยวิธีนี้ นิยามเชิงปฏิบัติว่า ข้อทดสอบจะไม่ลำเอียงเมื่อผู้สอบที่มีความสามารถเท่ากันมีโอกาสในการตอบข้อทดสอบถูก $(P(\theta))$ เท่ากัน ดังนั้น สมมติฐานทางสถิติของการทดสอบความลำเอียง คือ

$$H_0 : P_1(\theta) - P_2(\theta) = 0$$
$$H_1 : P_1(\theta) - P_2(\theta) \neq 0$$

อนึ่งการจะตรวจสอบว่าที่ทุก ๆ ค่าความสามารถ (θ) ผู้สอบในกลุ่มที่ 1 และกลุ่มที่ 2 จะมีโอกาสในการตอบถูกเท่ากัน ทำได้โดยการเปรียบเทียบโคน์ลักษณะของข้อทดสอบแต่ละข้อ สำหรับกลุ่ม 2 กลุ่ม โดยยอมให้มีความแตกต่างกันบ้าง วิธีการเปรียบเทียบ ICC ของกลุ่มผู้สอบ 2 กลุ่มเมื่ออยู่หลายวิธี ผู้วิจัยใช้ดัชนีที่คำนวณขึ้นที่ระหว่าง ICC โดยไม่บอกทิศทางของความลำเอียง ซึ่งคำนวณได้จากสูตรของ Osterlind (1983: 67) ซึ่งรายละเอียดได้กล่าวไว้แล้วในบทที่ 3

ความเหมือนและความแตกต่างของการวิเคราะห์หาความลำเอียง 3 วิธี

การวิเคราะห์หาความลำเอียงของข้อทดสอบ 3 วิธีดังกล่าวแล้ว มีความเหมือนและความแตกต่างกันในมิติต่าง ๆ ดังต่อไปนี้ (Osterlind, 1983: 78-79)

ตารางที่ 1: การเปรียบเทียบวิธีวิเคราะห์ความลำเอียงของข้อทดสอบ 3 วิธี

หัวข้อเรื่อง / วิธีวิเคราะห์	TID	χ^2 -test	ICC-3
1. ข้อตกลงเบื้องต้น	อันตรกิริยาระหว่างกลุ่มกับข้อทดสอบเป็นตัวบ่งชี้ถึงความลำเอียงภายในข้อทดสอบ	คะแนนรวมที่ได้จากแบบทดสอบเป็นตัวบ่งชี้ของความสามารถของผู้สอบ	แบบทดสอบต้องมีมิติเดียวและได้งลักษณะของข้อทดสอบสามารถแสดงฟังก์ชันของค่าความสามารถและโอกาสในการทำข้อทดสอบข้อนั้นๆ ได้
2. ดัชนีทำการตรวจสอบ	ค่าความยากง่าย (Δ)	ค่าความยากง่าย (p) จากการควบคุมการจำแนกของกลุ่ม	ค่าความยากง่าย (b) ค่าอำนาจจำแนก (a) และการเดา (c)
3. สิ่งที่ทำกรวิเคราะห์	อันตรกิริยาระหว่างการเป็นสมาชิกของกลุ่มกับการตอบข้อทดสอบถูก	ความแตกต่างในสัดส่วนของการตอบข้อทดสอบถูกเมื่อระดับของคะแนนรวมที่เท่ากัน	ความแตกต่างของโอกาสในการตอบข้อทดสอบถูก เมื่อระดับความสามารถที่เท่ากัน
4. เกณฑ์การตัดสินความลำเอียง	ระยะห่างจากจุดที่กำหนดไปยังเส้นแกนหลัก	ความมีนัยสำคัญทางสถิติของ χ^2 -test	ความแตกต่างของพื้นที่ระหว่าง ICC
5. การคำนวณและการแปลผล	ง่าย, ง่าย	ค่อนข้างง่าย, ง่าย	ยาก, ยาก

หัวข้อเรื่อง	วิธีวิเคราะห์	TID	χ^2 test	ICC-3
6. ข้อจำกัด	<ol style="list-style-type: none"> 1. นิยมใช้กับกลุ่มตัวอย่างขนาดเล็ก 2. ค่าความยากง่ายแปรเปลี่ยนตามกลุ่มผู้สอบ 3. คลาดเคลื่อนเมื่อข้อทดสอบมีค่าอำนาจจำแนกสูง 	<ol style="list-style-type: none"> 1. นิยมใช้กับกลุ่มตัวอย่างขนาดกลาง (= 100 คน/กลุ่ม) 2. ค่าความยากง่ายแปรเปลี่ยนตามกลุ่มผู้สอบ 3. ไม่มีกฎเกณฑ์ตายตัวในการแบ่งระดับคะแนน 4. ใช้คะแนนรวมเป็นตัวแทนของความสามารถ 5. วัตถุประสงค์กระจายของคะแนนและความแตกต่างของขนาดกลุ่มตัวอย่าง 	<ol style="list-style-type: none"> 1. ตัวอย่างมักมีขนาดใหญ่ (> 1,000 คน/กลุ่ม) 2. การประมาณค่าพารามิเตอร์ a และ c ทำให้ค่าคงที่ได้ยาก 3. ปัญหาการปรับสเกล ค่า a และ b 4. เสียค่าใช้จ่ายและเวลาในการใช้เครื่องคอมพิวเตอร์วิเคราะห์มาก 5. ผู้ใช้ต้องมีพื้นฐานความรู้ทางสถิติขั้นสูงมาก 6. สถิติที่ใช้ไม่สามารถหาได้โดยทั่ว ๆ ไป 	

หัวข้อ	วิธีวิเคราะห์	TID	χ^2 -test	ICC-3
7. ข้อดี		<ol style="list-style-type: none">1. คำนวณง่าย, ค่าใช้ จ่ายไม่สูง2. แปลผลและแสดงให้ผู้ อื่นที่ไม่มีพื้นฐานทาง สถิติเข้าใจได้ง่าย3. ไม่ต้องใช้กลุ่มตัว อย่างขนาดใหญ่	<ol style="list-style-type: none">1. คำนวณค่อนข้างง่าย กว่า ICC-32. มีเกณฑ์ตายตัวในการ แปลผล	<ol style="list-style-type: none">1. ให้รายละเอียดของ ความลำเอียงมาก2. ตรงตามหลักการวัดผล มากที่สุด3. ค่าพารามิเตอร์ไม่เปลี่ยน แปลงไปตามกลุ่ม4. ค่า a และ c เป็นจริง และเหมาะสมกับข้อมูล มากกว่าวิธีอื่น

III. งานวิจัยที่เกี่ยวข้องกับการวิเคราะห์หาความลำเอียง

1. งานวิจัยในต่างประเทศที่ใช้ข้อมูลจริง

ในปี ค.ศ. 1978 Rudner และ Convey (อ้างจาก Subkoviak, และคณะ, 1984: 51) ได้ทำการวิเคราะห์หาความลำเอียงของข้อทดสอบโดยใช้วิธีแปลงค่าความยากง่ายของข้อทดสอบ วิธีไคสแควร์ วิธีโค้งลักษณะของข้อทดสอบและวิธีวิเคราะห์องค์ประกอบ ข้อมูลในการวิจัยครั้งนี้เป็นข้อมูลจริงจากการใช้แบบทดสอบมาตรฐาน ผลการวิจัยปรากฏว่าวิธีโค้งลักษณะของข้อทดสอบ วิธีไคสแควร์ และวิธีแปลงค่าความยากง่ายเป็นวิธีที่เหมาะสม

ในปีต่อมา Ironson และ Subkoviak (1979) ได้ศึกษาถึงวิธีการในการวิเคราะห์หาความลำเอียงของข้อสอบ โดยใช้วิธีแปลงค่าความยากง่ายของข้อทดสอบ วิธีไคสแควร์ วิธีโค้งลักษณะของข้อทดสอบ และวิธีพอยท์ไบทีเรียล ผู้วิจัยได้นำแบบทดสอบมาตรฐานไปทดสอบกับคนอเมริกันผิวขาวและผิวดำ แล้วนำข้อมูลที่ได้มาทำการวิเคราะห์หาค่าสหสัมพันธ์ระหว่างวิธีการวิเคราะห์หาความลำเอียงแบบต่าง ๆ ผลการวิจัยพบว่าวิธีโค้งลักษณะของข้อทดสอบเป็นวิธีที่ดีที่สุด รองลงมาได้แก่ วิธีไคสแควร์ และวิธีแปลงค่าความยากง่าย ตามลำดับ

ในปีเดียวกัน Laksana (1979) ได้ประยุกต์ใช้วิธีโค้งลักษณะของข้อทดสอบที่มี 3 พารามิเตอร์และวิธีวิเคราะห์ความแปรปรวนในการวิเคราะห์แบบทดสอบ ชื่อ Iowa Tests of Basic Skills (ITBS) ในแบบทดสอบคำศัพท์และแบบทดสอบคณิตศาสตร์ระดับ 7 เพื่อหาความลำเอียงต่อเพศและผิว และการศึกษาพบว่าวิธีการวิเคราะห์ความแปรปรวนเป็นวิธีที่ใช้ได้ในทางปฏิบัติ แต่เป็นวิธีที่ไม่ครอบคลุมข้อมูลทุกลักษณะ วิธีโค้งลักษณะของข้อทดสอบเป็นวิธีที่ครอบคลุมมากกว่า แต่ไม่เหมาะสมในทางปฏิบัติเนื่องจากต้องใช้กลุ่มตัวอย่างขนาดใหญ่

นอกจากนี้ ในปีเดียวกันนั้น Intasuwana (1979) ได้ทำการวิเคราะห์หาความลำเอียงของแบบทดสอบเพื่อ International Association for the Evaluation of Educational Achievement (IEA) ด้วย 3 วิธี ได้แก่ วิธีโค้งลักษณะของข้อทดสอบที่มีพารามิเตอร์ 3 ตัว วิธี Rasch และวิธีไคสแควร์ กลุ่มตัวอย่างในการวิจัยในครั้งนี้เป็นวัยรุ่นชาวอังกฤษ อเมริกันและนิวซีแลนด์ ผลการวิจัยพบว่าความสัมพันธ์ระหว่างวิธีทั้ง 3 มีค่าอยู่ระหว่าง .51-.98 โดยวิธีวิเคราะห์แบบไคสแควร์และวิธี Rasch มีค่าสัมประสิทธิ์สหสัมพันธ์สูงถึง .98

ในปี ค.ศ. 1980 Sinnott (อ้างจาก Carton & Marco, 1982: 287) ได้ทำการวิเคราะห์ข้อมูลที่ได้จากแบบทดสอบ GMAT เพื่อจะแยกปัจจัยที่ทำให้ข้อทดสอบบางข้อง่ายหรือยากเกินไปสำหรับกลุ่มผู้สอบกลุ่มน้อย โดยใช้วิธีการกำหนดจุดค่าเดลต้า (Delta-Plot Method) แบบทดสอบ GMAT ที่ใช้ศึกษาทำการทดสอบในเดือนมกราคม ปี 1978 ประกอบด้วย 5 ส่วน คือ Reading Comprehension, Problem Solving, Practical Business Judgement, Data Sufficiency และ Usage ผู้วิจัยมุ่งศึกษาตัวแปรทางด้านสังคมของผู้สอบที่จะเป็นตัวทำให้เกิดความลำเอียง คือ เพศ เชื้อชาติ อายุ และความคล่องทางภาษา ตัวอย่างมีขนาดกลุ่มละ 400 คน ยกเว้น เชื้อชาติ Hispanic ที่มีจำนวนเพียง 309 คน ผู้วิจัยได้ทำการพัฒนาเกณฑ์ตัดสินความลำเอียง โดยการหาการกระจายของระยะห่างจากจุดตั้งฉากของค่าเดลต้า 2 กลุ่มที่เปรียบเทียบถึงเส้นแกนหลัก โดยใช้เกณฑ์การตัดสินความลำเอียง คือ $d = 0.7$ หลังจากที่พบข้อลำเอียงแล้ว ได้อภิปรายผลโดยมุ่งไปที่การวิเคราะห์เนื้อหาของข้อที่ลำเอียงเป็นสำคัญ และพบว่าวิธีดังกล่าวใช้ได้เป็นอย่างดี

ในปีต่อมา Stricker (อ้างจาก Carton & Marco, 1982: 289) ได้ศึกษาเปรียบเทียบดัชนีการตัดสินความลำเอียงของวิธีวิเคราะห์หาความลำเอียง 3 วิธี คือ 1) ระยะห่างซึ่งคำนวณโดยวิธีการกำหนดจุดค่าเดลต้า 2) ดัชนีโคสแควร์ที่ใช้ตัดสินความแตกต่างของค่าพารามิเตอร์ a , b และ c จากวิธีตอบตนเอง ข้อทดสอบ 3 พารามิเตอร์และ 3) ใช้ค่า Partial Correlation ระหว่างคะแนนข้อทดสอบแต่ละข้อของกลุ่มกับคะแนนรวมทั้งหมดของแบบทดสอบ GRE-AT (Graduate Record Examination: Aptitude Test ซึ่งทดสอบในเดือนตุลาคม ปี 1977 และเดือนมกราคม ปี 1978 โดยตัวแปรที่ศึกษาได้แก่เชื้อชาติ (ขาว, ดำ) และเพศ กลุ่มตัวอย่างประกอบด้วยกลุ่มผู้สอบชายผิวขาว (1,122 คน), หญิงผิวขาว (1,471 คน), ชายผิวดำ (284 คน) และหญิงผิวดำ (626 คน) โดยจับคู่วิเคราะห์ระหว่างกลุ่มหญิงผิวดำและขาวกับกลุ่มผู้สอบชายและหญิงผิวขาว และทำการวิเคราะห์ข้อมูลโดยแยกระหว่างผู้ชายและผู้หญิงในการวิเคราะห์ทางด้านเชื้อชาติ และวิเคราะห์แยกผู้สอบผิวดำและขาวในการวิเคราะห์เพศ ทั้งนี้เพื่อความคุมตัวแปรทั้ง 3 วิธี ผลการวิจัยพบว่าดัชนีที่ใช้สถิติโคสแควร์ตัดสินความแตกต่างระหว่างโค้งลักษณะของข้อทดสอบได้ผลดีกว่าอีก 2 วิธีที่ต้องตั้งเกณฑ์ เนื่องจากมีความคงที่และไม่เปลี่ยนแปลงตามเกณฑ์

ในปีเดียวกัน Shepard และคณะ (อ้างจาก Subkoviak, และคณะ, 1984: 50) ได้ทำการวิเคราะห์หาความลำเอียงของข้อทดสอบ 6 วิธี คือ



วิธีแปลงค่าความยากง่าย วิธีไคสแควร์ 2 วิธี วิธีโค้งลักษณะข้อทดสอบ 2 วิธี และวิธีพอยท์ไบท์เรียล แบบทดสอบที่ใช้เป็นแบบทดสอบมาตรฐาน สำหรับกลุ่ม ตัวอย่างของการวิจัยครั้งนี้เป็นชาวอเมริกัน ผิวขาว ผิวดำ และ Chicano ผลการวิจัยพบว่าวิธีโค้งลักษณะของข้อทดสอบที่มีพารามิเตอร์ 3 ตัว เป็นวิธีที่ดีที่สุด รองลงมาได้แก่หนึ่งในสองของวิธีวิเคราะห์ด้วยไคสแควร์ อย่างไรก็ตามวิธีแปลงค่าความยากง่ายของข้อทดสอบเป็นวิธีที่ใช้ได้ดีในทางปฏิบัติ ทั้งนี้เพราะวิธีแปลงค่าความยากง่ายของข้อทดสอบคำนวณได้ง่ายกว่าวิธีโค้งลักษณะข้อทดสอบที่มีพารามิเตอร์ ตัวเดียว และผลการวิเคราะห์ปรากฏว่าค่าสัมประสิทธิ์ระหว่างวิธีทั้งสองค่อนข้างสูง อนึ่งในปี ค.ศ. 1981 ปีเดียวกันนั้น Alderman และ Holland (อ้างจาก Carton and Marco 1982: 298) ได้ทำการวิจัยแบบทดสอบ TOEFL โดยมุ่งศึกษาตัวแปรทางด้านกลุ่มภาษา 7 กลุ่ม คือ Germanic, Spanish, African, Chinese I, Chinese II, Japanese และ Arabic ใช้ตัวอย่าง แต่ละกลุ่ม 500 ถึง 1,000 คน และใช้กลุ่มภาษาจีน 2 กลุ่มเป็นเกณฑ์ในการ ประเมินผลความลำเอียงและใช้วิธีวิเคราะห์ 2 วิธีคือ 1) วิธีวิเคราะห์โดย Log-Linear Model ซึ่งเป็นวิธีหนึ่งของวิธีไคสแควร์ และ 2) วิธีการกำหนด จุดเดลต้า (Delta-Plot Method) ผลการวิจัยพบว่าการวิเคราะห์โดยวิธี Log-Linear Model มีความไวมากเพราะพบข้อทดสอบที่มีความลำเอียงมากกว่าอีกวิธีหนึ่ง ดังนั้นผู้วิจัยจึงเสนอแนะว่าข้อทดสอบดังกล่าวควรต้องได้รับการนำไปพิจารณาอย่างลึกซึ้งโดยผู้เชี่ยวชาญทางภาษาและทำการวิเคราะห์ซ้ำเพื่อยืนยัน อีกครั้งหนึ่ง

ในปี ค.ศ. 1982 Ironson และ Craig (อ้างจาก Subkoviak และคณะ, 1984: 50) ได้ใช้วิธีวิเคราะห์หาความลำเอียงของข้อสอบ 5 วิธีคือ วิธีแปลงค่าความยากง่ายของข้อทดสอบ วิธีวิเคราะห์ด้วยไคสแควร์ 2 วิธี และวิธีโค้งลักษณะของข้อสอบ 2 วิธี ในการสร้างข้อสอบนั้น ผู้วิจัยตั้งใจที่จะสร้างข้อทดสอบใหม่มีความลำเอียงต่อผู้ชาย หลังจากนั้นผู้วิจัยได้ให้ผู้สอบทั้งชายและหญิง ประมาณค่า (Rate) ความลำเอียงของข้อสอบแต่ละข้อออกมาเป็น 5 ระดับ เมื่อได้รับข้อมูลมาแล้ว ผู้วิจัยจึงทำการวิเคราะห์หาความลำเอียงของข้อสอบแต่ละข้อด้วยวิธีการ 5 วิธีดังที่ได้กล่าวมา ผลการวิจัยพบว่าการประมาณค่าความลำเอียงของข้อสอบกับการวิเคราะห์หาความลำเอียงของข้อสอบทั้ง 5 วิธี มีความสัมพันธ์กันในระดับสูง นอกจากนี้ยังพบว่าค่าสัมประสิทธิ์สหสัมพันธ์ของจำนวนข้อทดสอบที่มีความลำเอียงจากวิธีการวิเคราะห์หาความลำเอียงของข้อสอบทั้ง 5 วิธีมีค่าสูงเช่นกัน

ในปี ค.ศ. 1984 Subkoviak และคณะ (1984) ได้ทำการเปรียบเทียบวิธีการวิเคราะห์หาความลำเอียงของข้อทดสอบ 3 วิธี เครื่องมือที่ใช้ในการวิจัยได้แก่แบบทดสอบคำศัพท์ชนิดเลือกตอบ 4 ตัวเลือก จำนวน 50 ข้อ แบบทดสอบนี้ประกอบด้วยคำศัพท์ภาษาอังกฤษมาตรฐาน (Standard English Vocabulary จำนวน 40 ข้อ และเป็นคำศัพท์เกี่ยวกับคำแสลง (Slang) ของคนผิวดำ 10 ข้อ ผู้วิจัยได้นำแบบทดสอบฉบับนี้ไปทดสอบกับนักเรียนผิวดำจำนวน 1,008 คน และนักเรียนผิวขาวจำนวน 1,021 คน ผลการวิจัยปรากฏว่า วิธีโค้งลักษณะของข้อทดสอบที่มีพารามิเตอร์ 3 ตัว เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาได้แก่ วิธีโคสแควร์ ส่วนวิธีแปลงค่าความยากง่าย หรือวิธีกำหนดจุดค่าเดลต้า เป็นวิธีที่มีข้อจำกัดแต่ก็สามารถนำไปใช้ในทางปฏิบัติได้ดีในกรณีที่ไม่มีเครื่องคอมพิวเตอร์

2. งานวิจัยในต่างประเทศที่ใช้ข้อมูลสำรวจ

ในปี ค.ศ. 1979 Rudner และคณะ (อ้างจาก Burrill, 1982: 162) ได้ศึกษาเปรียบเทียบวิธีการวิเคราะห์หาความลำเอียงของข้อทดสอบ โดยใช้ข้อมูลจำลอง (Simulate Data) 5 วิธี คือ 1) TID (Transformed Item Difficulties หรือ Delta-Plot Method) 2) ICC-3 (Three parameter Logistic Model) 3) ICC-1 (Rasch Model), 4) X^2 ที่แบ่งช่วงความสามารถเป็น 5 ช่วง และ 5) X^2 ที่แบ่งช่วงความสามารถหลายช่วง ผู้วิจัยได้สร้างข้อทดสอบที่มีความลำเอียงในพารามิเตอร์ A และ B โดยมีส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์เหล่านี้้อย่างละ 4 ค่า และใช้แบบทดสอบที่มีความยาว 4 ขนาด คือ 20, 40, 60 และ 80 ข้อ ใช้ผู้สอบ 2 กลุ่ม ๆ ละ 1,200 คน ผลการวิจัยพบว่าทุก ๆ วิธี ยกเว้นวิธีที่ 5 มีความสัมพันธ์กันเพิ่มขึ้นระหว่างการตรวจสอบความลำเอียงและการแยกความลำเอียงเมื่อความยากของแบบทดสอบเพิ่มขึ้น และเพิ่มมากที่สุดเมื่อแบบทดสอบมีความยาวอยู่ระหว่าง 20 และ 40 ข้อ วิธีที่ได้ผลดีที่สุดได้แก่วิธี ICC-3 รองลงมาได้แก่วิธี X^2 ที่แบ่งความสามารถเป็น 5 ช่วง ส่วนวิธีที่ได้ผลน้อยที่สุด ได้แก่วิธี ICC-1

อนึ่งยังพบว่าถ้าความลำเอียงไม่ได้เกิดจากที่พารามิเตอร์ B กล่าวคือถ้าความแตกต่างของกลุ่มขึ้นอยู่กับค่าอำนาจจำแนกของข้อกระเทง (พารามิเตอร์ A) แล้ว วิธี X^2 (5 ช่วง) มีประสิทธิภาพมากที่สุดในการตรวจสอบความลำเอียง รองลงมาคือวิธี ICC-1 และต่ำสุดคือวิธี TID

แต่ถ้าความล่าเอียงเกิดจากพารามิเตอร์ B เพียงตัวเดียวแล้ว วิธี TID และวิธี X^2 (5 ช่วง) มีประสิทธิภาพมากที่สุด วิธี X^2 และ TID ดังกล่าวสามารถตรวจสอบหาความล่าเอียงได้มากหรือน้อยขึ้นอยู่กับพารามิเตอร์ A

นอกจากนี้ยังพบว่าจำนวนข้อทดสอบที่มีความล่าเอียงที่พบจาก วิธี X^2 (5 ช่วง) และวิธี TID และระหว่างวิธี X^2 (5 ช่วง) กับวิธี X^2 (หลายช่วง) มีความสัมพันธ์กันสูงแต่ความสัมพันธ์ระหว่างวิธี ICC-1 กับวิธีอื่นค่อนข้างต่ำ

ดังนั้น Rudner และคณะ สรุปผลว่าวิธีตรวจสอบความล่าเอียง 3 วิธีที่ให้การประมาณที่ค่อนข้างตรงและแม่นยำ คือวิธี ICC-3 วิธี X^2 (5 ช่วง) และวิธี TID (Delta-Plot Method) วิธี ICC-3 จะให้ค่าแม่นยำมากน้อยเพียงใด ขึ้นอยู่กับการประมาณค่าพารามิเตอร์ว่าเที่ยงตรงหรือไม่ วิธี X^2 จะมีประสิทธิภาพมากหรือน้อยขึ้นอยู่กับการแบ่งช่วงการประมาณความสามารถว่าเหมาะสมเพียงใด ส่วนวิธี TID จะให้ค่าที่ไม่ไวต่อความล่าเอียงที่เกิดจากการจำแนกของกลุ่มต่างกัน

อนึ่ง ในปีเดียวกันนี้ Merz และ Grossen (อ้างจาก Subkoviak และคณะ, 1984: 51) ได้ใช้ทฤษฎีโค้งลักษณะของข้อสอบในการหาความล่าเอียงของข้อสอบโดยใช้ข้อมูลจำลองเช่นเดียวกัน โดยใช้วิธีแปลงค่าความยากง่ายของข้อทดสอบ วิธีโคลสแควร์ วิธีโค้งลักษณะของข้อทดสอบ 2 วิธี และวิธีวิเคราะห์ห้วงค์ประกอบ ผลการวิจัยปรากฏว่า วิธีแปลงค่าความยากง่ายของข้อทดสอบ เป็นวิธีการวิเคราะห์หาความล่าเอียงที่ดีที่สุด วิธีโค้งลักษณะของข้อทดสอบที่มีพารามิเตอร์ 3 ตัว และวิธีโคลสแควร์ต่างก็เป็นวิธีการที่ใช้ได้ดี

นอกจากนี้ ในปี ค.ศ. 1980 Rudner และคณะ (1980b) ได้ใช้ทฤษฎีโค้งลักษณะของข้อทดสอบในการวิเคราะห์หาความล่าเอียงของข้อสอบอีกครั้งหนึ่ง ในการวิจัยครั้งนี้ผู้วิจัยได้สร้างข้อมูลจำลองขึ้นใช้ ส่วนวิธีในการวิเคราะห์นั้นมี 7 วิธีดังนี้ คือ วิธีแปลงค่าความยากง่ายของข้อทดสอบ 2 วิธี วิธีโคลสแควร์ 2 วิธี และวิธีโค้งลักษณะของข้อทดสอบ 3 วิธี ผลการวิจัยปรากฏว่าวิธีโค้งลักษณะของข้อทดสอบที่มีพารามิเตอร์ 3 ตัว เป็นวิธีการหาความล่าเอียงของข้อสอบจากข้อมูลจำลองที่ดีที่สุด รองลงมาได้แก่หนึ่งในสองของวิธีการวิเคราะห์ด้วยโคลสแควร์ ส่วนวิธีแปลงค่าความยากง่ายของข้อทดสอบเป็นทางเลือกที่ผู้วิจัยแนะนำให้ใช้ได้ ในทางปฏิบัติเพราะคำนวณง่ายและเข้าใจง่าย

3. งานวิจัยในประเทศ

นับจนถึงปัจจุบันนี้ ปรากฏว่าเท่าที่ค้นพบมีงานวิจัยเกี่ยวกับการวิเคราะห์หาความลำเอียงของข้อทดสอบในประเทศไทย 3 เรื่อง กล่าวคือ

ในปี พ.ศ. 2526 ชัชชัย เผ่าพงษ์ (2526) ได้ศึกษาหาความลำเอียงของข้อทดสอบของแบบทดสอบมาตรฐานวัดความถนัดทางการเรียนด้านคณิตศาสตร์และภาษาในระดับมัธยมศึกษาตอนต้น จำนวน 2 ฉบับ ซึ่งพัฒนาโดยสำนักทดสอบทางการศึกษาและจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร ระหว่างกลุ่มผู้สอบเพศชายและหญิง เนื่องจากผู้วิจัยมีความสงสัยว่าแบบทดสอบวัดความถนัดทางการเรียนที่ได้รับการพัฒนาเป็นแบบทดสอบมาตรฐานแล้วนั้น สามารถวัดสิ่งที่ต้องการวัดได้จริงหรือไม่ เพราะว่าแบบทดสอบความถนัดหรือผลสัมฤทธิ์ทางการเรียน มักจะมีข้อตกลงว่าจำนวนข้อทดสอบทั้งฉบับมีข้อทดสอบที่ผู้สอบต่างก็มีความเคยชินและไม่เคยชินอยู่เท่า ๆ กัน ซึ่งในความเป็นจริงแล้วมีสภาพการณ์หลายอย่าง เช่น เพศ สภาพทางด้านภูมิศาสตร์ และอื่น ๆ ที่มีอิทธิพลบางอย่างทางด้านเศรษฐกิจและสังคม ทำให้ผู้สอบมีความแตกต่างกันในด้านประสบการณ์ จึงเกิดการเสียเปรียบหรือได้เปรียบทางด้านเนื้อหาของข้อทดสอบแตกต่างกัน กลุ่มตัวอย่างที่ใช้เป็นกลุ่มนักเรียนชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2524 เพศชายและหญิงจากทุกภาคภูมิศาสตร์ทั่วประเทศ คือ ภาคเหนือ ภาคกลาง ภาคใต้ ภาคตะวันออกและภาคตะวันออกเฉียงเหนือ ซึ่งเป็นการควบคุมตัวแปรทางสภาพภูมิศาสตร์ที่อาจจะเป็นตัวแปรที่จะทำให้เกิดความลำเอียงได้ ผู้วิจัยได้ใช้วิธีวิเคราะห์หาความลำเอียงของข้อทดสอบด้วย เครื่องลักษณะของข้อกระทงที่มีพารามิเตอร์ 3 ตัว ผลการวิจัยพบว่าแบบทดสอบวัดความถนัดทางการเรียนด้านคณิตศาสตร์มีข้อทดสอบที่ลำเอียงต่อกลุ่มนักเรียนชายและหญิงในระดับปานกลางขึ้นไป จำนวน 5 ข้อ เป็นข้อที่มีความลำเอียงต่อนักเรียนชายโดยเฉพาะ 4 ข้อ และเป็นข้อกระทงที่มีความลำเอียงต่อกลุ่มนักเรียนหญิงในช่วงความสามารถแรกและลำเอียงต่อกลุ่มนักเรียนชายในช่วงความสามารถต่อมาจำนวน 1 ข้อ ส่วนแบบทดสอบวัดความถนัดทางการเรียนด้านภาษาเกี่ยวกับการอ่านเข้าใจความจำนวน 40 ข้อนั้น ข้อทดสอบที่มีความลำเอียงในระดับปานกลางขึ้นไป มีจำนวน 9 ข้อ เป็นข้อที่มีความลำเอียงต่อกลุ่มนักเรียนชายโดยเฉพาะ 1 ข้อ และอีก 6 ข้อ มีความลำเอียงต่อกลุ่มนักเรียนหญิงโดยเฉพาะ และมี 2 ข้อ ซึ่งมีความลำเอียงต่อนักเรียนชาย ที่มีความสามารถต่ำ ดังนั้น ผู้วิจัยจึงสรุปว่าแบบทดสอบดังกล่าวมีความลำเอียง และวิธีการวิเคราะห์หาความลำเอียงด้วย เครื่องลักษณะของข้อทดสอบสามารถใช้ในการตรวจสอบดังกล่าวได้ผลถูกต้องดี

ต่อมาในปี พ.ศ.: 2530 กศนีย์ พีรมนตรี (2530: 77) ได้ทำการวิเคราะห์หาความลำเอียงของแบบทดสอบวิชาคณิตศาสตร์ระดับชั้นมัธยมศึกษาปีที่ 6 โดยการเปรียบเทียบจำนวนข้อทดสอบที่มีความลำเอียงระหว่างนักเรียนในกรุงเทพมหานครกับกลุ่มนักเรียนในภาคภูมิศาสตร์ทั้ง 5 ภาค ด้วยวิธีวิเคราะห์ 3 วิธี คือ 1) วิธีกำหนดจุดเดลต้า 2) วิธีทดสอบความแตกต่างระหว่างกลุ่มด้วย χ^2 -test 2 รูปแบบ และ 3) วิธีโค้งลักษณะของข้อทดสอบที่มีพารามิเตอร์ 3 ตัว หรือที่เรียกว่าวิธีการตอบสนองต่อข้อทดสอบ 3 พารามิเตอร์ และใช้กลุ่มตัวอย่างทั้งหมด 7,036 คน ผลการวิจัยพบว่า การวิเคราะห์หาความลำเอียงแต่ละวิธีได้ข้อทดสอบที่มีความลำเอียงจำนวนแตกต่างกัน แต่วิธีที่ 3 เป็นวิธีที่พบข้อทดสอบที่มีความลำเอียงจำนวนมากที่สุด และแต่ละวิธีพบข้อทดสอบที่ลำเอียงต่อภาคภูมิศาสตร์จำนวนแตกต่างกัน และบางข้อที่พบว่าลำเอียงก็แตกต่างกัน กล่าวคือ บางวิธีพบว่าข้อทดสอบบางข้อเป็นข้อที่มีความลำเอียง แต่บางวิธีกลับพบว่าข้อทดสอบนั้นไม่มีความลำเอียง ข้อทดสอบที่ลำเอียงที่การวิเคราะห์ทั้ง 3 วิธี ค้นพบเหมือนกัน โดยมากมักได้แก่ข้อทดสอบที่ง่ายสำหรับผู้สอบในกรุงเทพมหานคร จากผลการวิจัยครั้งนี้ผู้วิจัยได้เสนอแนะว่าในงานวิจัยที่มีความพร้อม และต้องการรายละเอียดของความลำเอียงมาก ควรใช้การวิเคราะห์หาความลำเอียงโดยวิธีที่ 3 ซึ่งได้แก่วิธีการตอบสนองต่อข้อทดสอบ 3 พารามิเตอร์ (วิธี ICC-3)

ต่อมาในปี พ.ศ. 2531 สรุศักดิ์ อมรรัตนศักดิ์ (2531: 199) ได้ทำการศึกษาเปรียบเทียบผลของวิธีวิเคราะห์หาความลำเอียงของข้อทดสอบ 4 วิธีคือ 1) วิธีวิเคราะห์ความแปรปรวน 2) วิธีแปลงค่าความยากง่ายของข้อทดสอบ 3) วิธีโค้งลักษณะของข้อทดสอบที่มีพารามิเตอร์ 1 ตัว และ 4) วิธีโค้งลักษณะของข้อทดสอบที่มีพารามิเตอร์ 3 ตัว เมื่อใช้วิเคราะห์หาความลำเอียงต่อเพศของข้อทดสอบที่ใช้สอบแข่งขันเพื่อบรรจุเป็นข้าราชการ 4 ฉบับ และใช้ผลวิจัยเพศชาย 1,170 คน และเพศหญิงอีก 1,170 คน ผลการวิจัยพบว่าวิธีวิเคราะห์หาความลำเอียงแต่ละวิธีพบข้อทดสอบที่มีความลำเอียงต่างกัน โดยวิธีโค้งลักษณะของข้อทดสอบที่มีพารามิเตอร์ 3 ตัว พบข้อทดสอบที่มีความลำเอียงจำนวนมากที่สุด รองลงมาได้แก่ วิธีวิเคราะห์ความแปรปรวน ส่วนวิธีแปลงค่าความยากง่ายของข้อทดสอบพบข้อทดสอบที่มีความลำเอียงจำนวนน้อยที่สุด และยังพบว่าค่าสัมประสิทธิ์สหสัมพันธ์ของดัชนีที่บ่งบอกความลำเอียงของข้อทดสอบของทั้ง 4 วิธีสูงมาก คือ มีค่า r_{xy} ระหว่าง 0.754-0.992 แต่อย่างไรก็ตาม ผู้วิจัยกล่าวว่าวิธีการวิเคราะห์ทั้ง 4 วิธี ยังไม่อาจกล่าวได้ว่า วิธีใดเป็นวิธีที่ดีที่สุด เพราะบางวิธีผู้วิจัยเป็นผู้กำหนดเกณฑ์ตัดสินความลำเอียงเอง ดังนั้นหากว่าเกณฑ์เปลี่ยนไปจำนวนข้อ

ทดสอบที่มีความลำเอียงก็เปลี่ยนไปด้วย และบางวิธีมีข้อจำกัดมากจึงเสนอแนะว่า
วิธีที่ควรนำมาใช้ ได้สะดวกคือวิธีวิเคราะห์ความแปรปรวน

ดังนั้น จากผลการวิจัยที่เกี่ยวข้องต่าง ๆ ดังกล่าวมาแล้ว ปรากฏว่าข้อ
ค้นพบที่ได้มีทั้งที่สอดคล้องและ ไม่สอดคล้องกัน กล่าวคือสำหรับข้อมูลจำลอง
(Simulate Data) Rudner และคณะพบว่าวิธีใดึ่งลักษณะของข้อทดสอบที่มี
พารามิเตอร์ 3 ตัว เป็นวิธีที่ดีที่สุด ส่วน Merz และ Grossen กลับพบว่าวิธี
แปลงค่าความยากง่ายของข้อทดสอบเป็นวิธีที่ดีที่สุด แต่สำหรับข้อมูลจริง (Real
Data) นั้น ปรากฏว่าข้อค้นพบที่ได้ส่วนใหญ่ค่อนข้างจะสอดคล้องกัน คือต่างก็
พบว่าวิธีใดึ่งลักษณะของข้อทดสอบเป็นวิธีที่เหมาะสมในทางทฤษฎี ถูกต้อง และ
ตรงที่สุด ในการวิเคราะห์หาความลำเอียงของข้อสอบ แต่อย่างไรก็ตาม วิธีนี้
เป็นวิธีที่มีความยุ่งยากและซับซ้อนทางคณิตศาสตร์ ต้องใช้กลุ่มตัวอย่างขนาดใหญ่
และข้อทดสอบจำนวนมาก อีกทั้งเสียค่าใช้จ่ายในการคำนวณด้วยคอมพิวเตอร์มาก
จึงเป็นสาเหตุส่วนหนึ่งที่ทำให้ผู้วิจัยสมัยก่อนให้ใช้วิธีการอื่น ในการวิเคราะห์หา
ความลำเอียงของข้อสอบแม้จะไม่เหมาะสมในเชิงทฤษฎีมากนัก แต่ก็เห็นว่าเป็น
วิธีการที่ใช้ได้ในทางปฏิบัติ วิธีที่ได้รับการสนับสนุนให้ใช้วิเคราะห์หาความ
ลำเอียงของข้อสอบแทนคือวิธีวิเคราะห์ความแปรปรวน

จากข้อค้นพบดังกล่าวทำให้ไม่อาจสรุปข้อความรู้ที่แน่นอนเกี่ยวกับวิธีการ
วิเคราะห์หาความลำเอียงของข้อสอบได้ และเพื่อเป็นการตรวจสอบ (Verify)
ข้อความรู้ดังกล่าวอีกครั้งหนึ่ง ผู้วิจัยจึงต้องการที่จะเปรียบเทียบผลการวิเคราะห์
หาความลำเอียงของข้อทดสอบวิธีต่าง ๆ อีกครั้งหนึ่ง โดยใช้ข้อทดสอบภาษา
อังกฤษเพื่อเข้ามหาวิทยาลัย 6 ฉบับ เป็นเครื่องมือ เนื่องจากเป็นแบบทดสอบที่
เกี่ยวข้องกับภาคตัดสินอนาคตของนักเรียนจำนวนมากที่สุดในประเทศไทย จึงนำ
จะเป็นแบบทดสอบที่มีคุณภาพดีที่สุด และปราศจากความลำเอียงต่อเพศหรือภาคภูมิ-
ศาสตร์ของผู้สอบ มิฉะนั้นแล้วแบบทดสอบดังกล่าวอาจก่อให้เกิดผลกระทบทางลบแก่
ผู้สอบที่ได้รับความลำเอียงทางลบจากแบบทดสอบนั้น ซึ่งเป็นเรื่องที่คณะกรรมการ
สร้างข้อสอบ คงไม่ปรารถนาเป็นอย่างยิ่ง เพราะผลจะทำให้เกิดความผิดพลาดใน
การตัดสินผลการสอบตามไปด้วย และมีผลต่อคุณภาพของบัณฑิตรวมทั้งต่อคุณภาพ
ของพลเมืองไทยด้วย และที่สำคัญมากที่สุดก็คือผลทางลบต่อชะตากรรมของผู้สอบ
เข้ามหาวิทยาลัยไม่ได้เป็นจำนวนมากด้วย

บทที่ 3 วิธีดำเนินการวิจัย

เพื่อให้บรรลุวัตถุประสงค์ของการวิจัยดังกล่าวมาแล้ว ผู้วิจัยได้ดำเนินการวิจัยดังต่อไปนี้

ประชากร (populations)

การวิจัยครั้งนี้มีประชากร 6 กลุ่ม คือผู้สมัครสอบคัดเลือกเข้ามหาวิทยาลัย ปีการศึกษา พ.ศ. 2531-2533 ที่สอบแบบทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย ชุด กข. และ/หรือ ชุด กขค. ซึ่งมีจำนวนแตกต่างกัน บุคคลเหล่านี้เป็นผู้สอบเพศชาย หรือหญิง และมาจากภาคภูมิศาสตร์ที่แตกต่างกันและมีจำนวนแตกต่างกันในแต่ละปี ประชากรเหล่านี้ผู้วิจัยได้จากแถบแม่เหล็กบันทึกข้อมูลคอมพิวเตอร์ของฝ่ายวิจัยสถาบันและสารสนเทศของทบวงมหาวิทยาลัย ซึ่งอนุญาตให้ใช้เพื่อการวิจัยได้

ผลวิจัย (subjects)

ผลวิจัยของการวิจัยครั้งนี้ ได้แก่ตัวอย่างของประชากรดังกล่าวแล้ว ซึ่งผู้วิจัยได้ใช้เครื่องคอมพิวเตอร์สุ่มแบบง่าย (simple random sampling technique) มาใช้เพื่อการวิจัย ตามสัดส่วนที่แตกต่างกันตามจำนวนมากหรือน้อยของประชากรในแต่ละกลุ่ม โดยให้ผลวิจัยจากแต่ละภาคภูมิศาสตร์ มีจำนวนเพศชายและหญิงเท่ากัน (เพื่อป้องกันความลำเอียงที่เกิดจากเพศต่างกัน) และรวมกันแล้ว จำนวนประมาณ 1,500 คน ทั้งนี้เพื่อให้มีผลวิจัยเมื่อจำแนกเป็นเพศชายและหญิงแล้วมีจำนวนอย่างละประมาณ 3,000 คน ซึ่งจำนวนผลวิจัยประมาณ 1,500-3,000 คนนี้ เป็นจำนวนที่เพียงพอสำหรับการวิเคราะห์ข้อทดสอบแบบ Logistic Model เพื่อให้ได้ค่าพารามิเตอร์ที่คงที่ (Wingersky, 1982: 70) และเป็นจำนวนสูงสุดที่สามารถใช้ได้กับโปรแกรมวิเคราะห์ข้อสอบแบบ Classical Model (สัทธัมภ์ สุกมลสันต์, 2533: 74)

อนึ่ง หากถือว่าการสอบเข้ามหาวิทยาลัยในแต่ละปีมีลักษณะได้หรือตก (สอบเข้าได้หรือสอบเข้าไม่ได้) แล้ว จำนวนกลุ่มตัวอย่างหรือผลวิจัยที่เหมาะสมเพื่อการวิจัย เมื่อกำหนดให้ความเชื่อมั่นในการสุ่มตัวอย่างง่ายเท่ากับร้อยละ 99 และความคลาดเคลื่อนในการสุ่มไม่เกินร้อยละ 5 แล้ว ปรากฏว่าได้กลุ่มตัวอย่างของประชากรแต่ละกลุ่มน้อยกว่าผลวิจัยที่ใช้ในการวิจัยครั้งนี้ ดังข้างจึงเชื่อได้ว่าจำนวนผลวิจัยของประชากรแต่ละกลุ่มมีจำนวนเพียงพอ และเป็นตัวแทนที่ดีของประชากรสำหรับการวิจัยครั้งนี้ (Yamane, 1967: 398)

ประชากรและผลวิจัยของแต่ละกลุ่มมีดังนี้

ตารางที่ 2: จำนวนประชากรและผลวิจัยของผู้สมัครสอบภาษาอังกฤษ
เข้ามหาวิทยาลัย ชด กท. ปี พ.ศ. 2531-2533

ปีการศึกษา	ภาค/เพศ	ประชากร	ผลวิจัย**	ร้อยละ	ผลวิจัยที่เหมาะสม*	
					จำนวน	ร้อยละ
2531 N = 62,821	กลาง	40,034	1,500	3.75	881	2.20
	ตะวันออก	2,296	1,500	65.33	662	28.83
	ตะวันออกเฉียงเหนือ	9,481	1,500	15.82	826	8.71
	เหนือ	5,368	1,500	27.94	783	14.59
	ใต้	5,642	1,500	26.59	783	13.88
	ชาย	29,189	3,000	10.28	881	3.02
	หญิง	33,632	3,000	8.92	881	2.62
2532 N = 66,155	กลาง	40,250	1,500	3.73	881	2.19
	ตะวันออก	2,671	1,500	56.16	662	24.78
	ตะวันออกเฉียงเหนือ	10,721	1,500	13.99	826	7.70
	เหนือ	6,209	1,500	24.16	783	12.61
	ใต้	6,304	1,500	23.79	783	12.42
	ชาย	30,445	3,000	9.85	881	2.89
	หญิง	35,710	3,000	8.40	881	2.47
2533 N = 78,580	กลาง	47,679	1,500	3.15	881	1.85
	ตะวันออก	2,843	1,500	52.76	692	24.34
	ตะวันออกเฉียงเหนือ	13,035	1,500	11.51	849	6.51
	เหนือ	7,601	1,500	19.74	809	10.64
	ใต้	7,422	1,500	20.21	809	10.90
	ชาย	36,672	3,000	8.18	881	2.40
	หญิง	41,908	3,000	7.16	881	2.10
	รวม	207,556	22,500	10.84	900	0.43

* ความเชื่อมั่น 99% และความคลาดเคลื่อน $\pm 5\%$

** จำนวนชายและหญิงในแต่ละภาคภูมิศาสตร์เท่ากัน

ตารางที่ 3: จำนวนประชากรและผลวิจัยของผู้สมัครสอบภาษาอังกฤษ
เข้ามหาวิทยาลัย ชุด กขค. ปี พ.ศ. 2531-2533

ปีการศึกษา	ภาค/เพศ	ประชากร	ผลวิจัย**	ร้อยละ	ผลวิจัยที่เหมาะสม*	
					จำนวน	ร้อยละ
2531 N = 27,259	กลาง	19,284	1,500	7.78	861	4.46
	ตะวันออก	921	424	46.04	474	51.46
	ตะวันออกเฉียงเหนือ	2,405	1,170	48.65	662	27.53
	เหนือ	2,054	1,052	51.22	621	30.23
	ใต้	2,595	1,334	51.41	662	25.51
	ชาย	7,777	2,740	35.23	809	10.40
	หญิง	19,482	2,740	14.06	861	4.42
2532 N = 27,566	กลาง	19,202	1,500	7.81	861	4.48
	ตะวันออก	975	494	50.67	474	48.62
	ตะวันออกเฉียงเหนือ	2,418	1,176	48.64	662	27.38
	เหนือ	2,160	1,180	54.68	621	28.75
	ใต้	2,811	1,392	49.52	692	24.62
	ชาย	7,702	2,874	37.31	809	10.50
	หญิง	19,864	2,874	14.47	861	4.33
2533 N = 30,673	กลาง	21,816	1,500	6.88	861	3.95
	ตะวันออก	989	484	48.94	474	47.93
	ตะวันออกเฉียงเหนือ	2,563	1,354	52.83	662	25.83
	เหนือ	2,571	1,500	58.34	662	25.75
	ใต้	2,734	1,500	54.86	692	25.31
	ชาย	8,660	3,000	34.64	818	9.45
	หญิง	22,013	3,000	13.63	869	3.95
	รวม	85,498	17,560	20.54	892	1.04

* ความเชื่อมั่น 99% และความคลาดเคลื่อน \pm 5%

** จำนวนชายและหญิงในแต่ละภาคภูมิศาสตร์เท่ากัน

เครื่องมือที่ใช้เพื่อการวิจัย

- การวิจัยครั้งนี้ใช้ข้อทดสอบจำนวน 600 ข้อของแบบทดสอบ 6 ชุดเป็นเครื่องมือ คือ
- 1-3 แบบทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย ชุด กข. สำหรับปีการศึกษา 2531-2533 ซึ่งเป็นแบบทดสอบที่สร้างขึ้น โดยคณะกรรมการสร้างแบบทดสอบของทบวงมหาวิทยาลัย แบบทดสอบนี้เป็นแบบทดสอบปรนัย (objective test) แบบเลือกตอบ (multiple-choice type) ที่มีตัวเลือก 4 ตัวเลือก และมีข้อทดสอบ 100 ข้อ
 - 4-6 แบบทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย ชุด กขค. สำหรับปีการศึกษา 2531-2533 ซึ่งมีลักษณะเหมือนกับแบบทดสอบชุดแรก

การเก็บรวบรวมข้อมูล

ผู้วิจัยได้ดำเนินการเก็บรวบรวมข้อมูลที่ใช้ในการวิจัยครั้งนี้ ดังนี้

1. เขียนเจตนาติดต่ออย่างเป็นทางการ เพื่อขอยืมแถบบันทึกข้อมูลคอมพิวเตอร์ (computer tape) ที่บันทึกผลการสอบรายวิชา 05 (ภาษาอังกฤษ ชุด กข.) และ 10 (ภาษาอังกฤษ ชุด กขค.) เป็นรายชื่อของผู้เข้าสอบทั้งหมด ประจำปีการศึกษา 2531-2533 จากทบวงมหาวิทยาลัย
2. นำแถบบันทึกข้อมูลดังกล่าวมาดำเนินการหาขนาดของประชากรทั้งหมดในแต่ละปี และแยกประเภทผู้เข้าสอบเป็นกลุ่มตามภาคภูมิศาสตร์ และเพศ

การวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลสำหรับการวิจัยนี้ อาศัยโปรแกรม SPSSX (The Social Packages for the Social Sciences, Version X) ของสถาบันบริการคอมพิวเตอร์ และโปรแกรมอื่น ๆ อีก ดังนี้

1. ใช้ Sub-program ชื่อ Sample ทำการสุ่มพลวิจัยที่แยกประเภทแล้ว ตามอัตราส่วนที่แตกต่างกันตามจำนวนมากหรือน้อยของประชากรในแต่ละกลุ่ม โดยกำหนดให้พลวิจัยในแต่ละกลุ่มภาคภูมิศาสตร์มีขนาดประมาณ 1,500 คน และมีเพศชายและหญิงเท่ากัน และให้พลวิจัยที่แยกเป็นกลุ่มเพศชายหรือเพศหญิงมีขนาดไม่เกิน 3,000 คน และมาจากทุก ๆ ภาคภูมิศาสตร์
2. ใช้ Sub-program ชื่อ Crosstab ตรวจสอบขนาดพลวิจัยว่ามีจำนวนเท่าใดในแต่ละกลุ่ม มาจากภาคใด และมีเพศใด เป็นต้น



← ทั้งนี้ เพื่อป้องกันความลำเอียงที่อาจเกิดจากเพศ หรือภาคภูมิศาสตร์ของ
พลวิจัยและเพื่อให้ได้ตัวแทนที่ดีของประชากรในแต่ละกลุ่มและแต่ละปี

3. ทำการวิเคราะห์หาความลำเอียงของข้อทดสอบวิธีที่ 1 คือ วิธีกำหนดจุดค่าเดลต้า
(Delta-Plot Method) ดังนี้

3.1 ใช้โปรแกรมชื่อ ITEMX ทำการวิเคราะห์ข้อทดสอบรายข้อตามวิธี
ประเพณีนิยม (Classical Model) เพื่อคำนวณหาค่าความยากง่าย
(Δ) และค่าอำนาจจำแนกของข้อทดสอบรายข้อและค่าสถิติต่าง ๆ ของ
แบบทดสอบ เช่น ค่าความเที่ยงแบบต่าง ๆ เป็นต้น ทั้งนี้สำหรับผู้สอบ
แต่ละกลุ่มและแบบทดสอบแต่ละฉบับ (ดูภาคผนวก ค.)

3.2 ใช้โปรแกรมชื่อ BIAS ซึ่งเป็นโปรแกรมที่ผู้วิจัยเขียนขึ้น เพื่อใช้กับเครื่อง
ไมโครคอมพิวเตอร์ ทำการคำนวณหาความลำเอียงของข้อทดสอบแต่ละข้อ
สำหรับผู้สอบแต่ละกลุ่มในแต่ละปีการศึกษา โดยอาศัยค่าความยากง่าย
(Δ) ที่ได้จากการใช้โปรแกรมชื่อ ITEMX โดยอาศัยสูตรต่อไปนี้
(Osterlind, 1983: 34) คือ

$$d_i = \frac{bx_i + a - y_i}{\sqrt{b^2 + 1}}$$

$$y = bx + a \quad (\text{สมการถดถอยเชิงเส้นตรง})$$

$$a = M_y - bM_x$$

$$b = \frac{(\sigma_y^2 - \sigma_x^2) \pm \sqrt{(\sigma_y^2 - \sigma_x^2)^2 + 4r_{xy}^2 \sigma_x^2 \sigma_y^2}}{2r_{xy} \sigma_x \sigma_y}$$

ในเมื่อ $d =$ ค่าระยะห่างตั้งฉากจากจุดอื่นดับค่าเดลต้าของข้อทดสอบไปยัง
เส้นแกนหลัก

X และ Y = ค่าเดลต้าแต่ละค่าของกลุ่มที่ 1 และ 2

a = ค่าที่ตัดแกน Y

b = ค่าความชันของ slope

$M_x =$ ค่าเฉลี่ยเลขคณิตของค่าเดลต้าของกลุ่มที่ 1

$M_y =$ ค่าเฉลี่ยเลขคณิตของค่าเดลต้าของกลุ่มที่ 2

- σ_x = ค่าเบี่ยงเบนมาตรฐานของค่าเฉลี่ยของกลุ่มที่ 1
- σ_y = ค่าเบี่ยงเบนมาตรฐานของค่าเฉลี่ยของกลุ่มที่ 2
- r_{xy} = ค่าสัมประสิทธิ์สหสัมพันธ์ของค่าเฉลี่ยของกลุ่มที่ 1 และ 2

3.3 คำนวณหาค่าเบี่ยงเบนมาตรฐานของระยะห่างตั้งฉาก จากสูตร

$$S_d = \sqrt{1 - r_{xy}^2}$$

4. ทำการวิเคราะห์หาความลำเอียงของข้อทดสอบวิธีที่ 2 คือวิธีทดสอบความแตกต่างของสัดส่วนของแต่ละกลุ่มด้วยไคสแควร์ (Chi-Square Method) ด้วยวิธี Marascuilo Method One (Osterlind, 1983: 49) ดังนี้

4.1 แบ่งผู้สอบแต่ละกลุ่มออกเป็น 3 กลุ่มย่อยตามระดับคะแนนรวมผลการสอบ คือ

0 - 40 คะแนน = กลุ่มความสามารถระดับต่ำ (กลุ่มต่ำ)

41 - 70 คะแนน = กลุ่มความสามารถระดับปานกลาง (กลุ่มปานกลาง)

71 - 100 คะแนน = กลุ่มความสามารถระดับสูง (กลุ่มสูง)

4.2 ทำการทดสอบความแตกต่างของสัดส่วนของผู้ตอบถูกหรือในการตอบผิด ข้อทดสอบแต่ละข้อของผู้สอบแต่ละระดับความสามารถ ครั้งละ 2 กลุ่ม (เช่น เพศชายและหญิง) ด้วย χ^2 test โดยการใช Sub-program ชื่อ Crosstab (ดูภาคผนวก ค.)

4.3 ทดสอบหาความลำเอียงของข้อทดสอบแต่ละข้อ โดยอาศัยค่า χ^2 ที่ระดับความมีนัยสำคัญ ได้รับการปรับค่าความคลาดเคลื่อนชนิดที่ 1 (Type I (α) error) แล้วคือ

$$\alpha = \frac{1}{j} \alpha_t$$

ในเมื่อ $\alpha_t = 0.05$

j = จำนวนกลุ่มระดับความสามารถ

ดังนั้นเกณฑ์ในการตัดสินความลำเอียงคือ $\chi^2_{1, 0.017} = 5.7242$

5. ทำการวิเคราะห์หาความลำเอียงของข้อทดสอบวิธีที่ 3 คือวิธีการตอบสนองของข้อทดสอบชนิด 3 พารามิเตอร์ (Three-Parameter Logistic Method) ดังนี้ (Osterlind, 1983: 60-69)

5.1 ทดสอบความเป็นมิติเดียว (unidimensionality) ของข้อทดสอบของแบบทดสอบแต่ละฉบับ ด้วยวิธีการวิเคราะห์ตัวประกอบ (Factor

- 5.1 ทดสอบความเป็นมิติเดียว (unidimensionality) ของข้อทดสอบของแบบทดสอบแต่ละฉบับ ด้วยวิธีการวิเคราะห์ตัวประกอบ (Factor Analysis) โดยใช้ Sub-program ของ SPSSX ชื่อ Factor ตามวิธีการของ Lord (1980: 21) ด้วยการวิเคราะห์ตัวประกอบสำคัญ (principal component) แล้วหมุนแกนด้วยวิธี Varimax และพบว่า eigenvalue ของตัวประกอบที่ 1 ของแบบทดสอบแต่ละฉบับมีขนาดแตกต่างจาก eigenvalue ของตัวประกอบที่ 2 มาก และ eigenvalue ของตัวประกอบที่ 2 มีขนาดใกล้เคียงกับ eigenvalue ตัวต่อ ๆ ไป จึงอาจสรุปได้ว่า แบบทดสอบแต่ละฉบับได้วัดเพียงมิติเดียว (ดูภาคผนวก ง.)
- 5.2 ใช้โปรแกรมชื่อ LOGIST ทำการวิเคราะห์ข้อทดสอบรายข้อตามวิธีโลจิสต์ (Logistic Model) เพื่อคำนวณหาค่าความยากง่าย (b) ค่าอำนาจจำแนก (a) และค่าการเดา (c) ของข้อทดสอบรายข้อ และค่าความสามารถของผู้สอบแต่ละคน (θ) ทั้งนี้สำหรับผู้สอบแต่ละกลุ่มและแบบทดสอบแต่ละฉบับ
- 5.3 ปรับเทียบ (equate) ค่าพารามิเตอร์ของผู้สอบแบบทดสอบแต่ละฉบับจากภาคภูมิศาสตร์ต่าง ๆ ให้อยู่ในมาตราวัด (scale) เดียวกันโดยใช้ค่าพารามิเตอร์ของผู้สอบจากภาคกลางเป็นหลัก และปรับเทียบค่าพารามิเตอร์ของผู้สอบภาคอื่นๆ โดยใช้ค่าพารามิเตอร์ของเพศชายเป็นหลัก โดยใช้สมการต่อไปนี้

ใช้ $y = ax + b$ (สมการถดถอยเชิงเส้นตรง) เปรียบเทียบ

$$a^*_{12} = (1/A)a_{12} \quad \text{สำหรับข้อที่ } i$$

$$b^*_{12} = Ab_{12} + B$$

$$A = \sigma^2_{bi1} / \sigma^2_{bi2}$$

$$B = b_{.1} - Ab_{.2}$$

ในเมื่อ $A =$ ค่าอำนาจจำแนกของข้อทดสอบก่อนปรับเทียบ

$B =$ ค่าความยากง่ายของข้อทดสอบก่อนปรับเทียบ

$\sigma^2_{bi1} =$ ค่าเบี่ยงเบนมาตรฐานของค่า B ของกลุ่มที่ใช้ในเป็นหลัก

σ_{bi2} = ค่าเบี่ยงเบนมาตรฐานของค่า B ของกลุ่มที่ต้องการปรับเทียบค่า

$b_{.1}$ = ค่าเฉลี่ยเลขคณิตของค่า B ของกลุ่มที่ใช้เป็นหลัก

$b_{.2}$ = ค่าเฉลี่ยเลขคณิตของค่า B ของกลุ่มที่ต้องการปรับเทียบค่า

5.4 คำนวณหาโค้งลักษณะของข้อทดสอบหรือ ICC (Item Characteristic Curve) ของข้อทดสอบแต่ละข้อของผู้สอบที่มีความสามารถเท่ากัน แต่อยู่ต่างกลุ่ม ซึ่งได้แก่ โอกาสในการตอบข้อทดสอบแต่ละข้อถูกต้องของผู้สอบที่มีความสามารถเท่ากันแต่อยู่ในกลุ่มต่างกัน ด้วยการใส่ค่าพารามิเตอร์ของกลุ่มหลักและค่าพารามิเตอร์ที่ได้ปรับเทียบค่าแล้ว โดยใช้สูตรต่อไปนี้ (Hambleton and Cook, 1977: 82):

$$P_1(\theta) = c_1 + (1-c_1) \frac{e^{Da_1(\theta - b_1)}}{1 + e^{Da_1(\theta - b_1)}}$$

ในเมื่อ a_1 = ค่าอำนาจจำแนกของข้อสอบข้อที่ i

b_1 = ค่าความยากง่ายของข้อสอบข้อที่ i

c_1 = ค่าการเดาของข้อสอบข้อที่ i

θ = ค่าความสามารถของผู้สอบ

D = ค่า Scale factor มีค่า 1.7

5.5 เปรียบเทียบความแตกต่างของพื้นที่ภายใต้โค้งลักษณะของข้อทดสอบ (ICC) ของผู้สอบต่างกลุ่มกัน แต่มีความสามารถเท่ากัน โดยสูตรต่อไปนี้ (Osterlind, 1983: 67)

$$\phi_i = \int_{-5}^5 [P(U_i = 1/\theta_j) - P(U_i = 1/\theta_j)] \Delta\theta$$

หรือ

$$\phi_i = \int_{-5}^5 P_1(\theta) \Delta\theta - \int_{-5}^5 P_2(\theta) \Delta\theta$$

ในเมื่อ $\Delta\theta = 0.005$

$\int P_1(\theta)$, $P(U_i = 1/\theta_j)$ = ค่าพื้นที่ของกลุ่มหลัก

$\int P_2(\theta)$, $P(U_j = 1/\theta_j)$ = ค่าพื้นที่ของกลุ่มที่เทียบปรับแล้ว

← อนึ่งการคำนวณในขั้นที่ 5.3-5.5 ใช้โปรแกรมคอมพิวเตอร์ชื่อ ICCTEST ที่ผู้วิจัยเขียนขึ้นใช้กับเครื่องไมโครคอมพิวเตอร์

6. เปรียบเทียบจำนวนข้อทดสอบที่มีความลำเอียงต่อเพศหรือภาคภูมิศาสตร์ของผู้สอบที่ได้จากการวิเคราะห์หาความลำเอียงวิธีที่ 1-3
7. ทดสอบหาความแตกต่างของจำนวนข้อทดสอบที่ลำเอียงจากการวิเคราะห์ทั้ง 3 วิธี รวมทั้งคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ของจำนวนข้อทดสอบที่ลำเอียง (generalize) ผลของการศึกษาดังกล่าว

สถิติที่ใช้

1. ใช้สถิติบรรยาย (descriptive statistics) เพื่อหาค่า \bar{x} , S.D. และความถี่ของข้อมูลต่าง ๆ ที่ต้องการบรรยายลักษณะทั่วไปของข้อมูล
2. ใช้การสุ่มแบบง่าย (Simple random sampling technique) เพื่อสุ่มผลวิจัย
3. ใช้ χ^2 test เพื่อทดสอบความลำเอียงวิธีที่ 2
4. ใช้ Factor Analysis เพื่อทดสอบความมีมิติเดียวของแบบทดสอบ
5. ใช้ Pearson Product-moment Correlation เพื่อหาความสัมพันธ์ของจำนวนข้อทดสอบที่ลำเอียงเมื่อวิเคราะห์ด้วยวิธีต่างกัน
6. ใช้สถิติต่าง ๆ ที่เกี่ยวข้องกับ การวิเคราะห์ข้อทดสอบรายข้อตามวิธีประเพณีนิยม (Classical Model) และวิธีโลจิส (Logistic Model) เช่น P , Δ , r_{b1s} และ KR_{20} เป็นต้น
7. ใช้สถิติอื่น ๆ เพื่อคำนวณหาระยะห่างตั้งฉากของคู่อันดับเดลต้า (d_1) และค่าความแตกต่างของฟังก์ชันได้โค้งลักษณะของข้อทดสอบ (ϕ_1) โดยอาศัยโปรแกรมคอมพิวเตอร์ที่ผู้วิจัยเขียนขึ้นใช้เอง (ดูภาคผนวก ง.)

จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 4

ผลการวิเคราะห์ข้อมูล

ผลการวิเคราะห์ข้อมูลของการวิจัยนี้แบ่งออกได้เป็น 2 ส่วนใหญ่ ๆ และแต่ละส่วนประกอบด้วยส่วนย่อย ๆ อีกดังต่อไปนี้

- ก. ความลำเอียงต่อเพศและต่อภาคภูมิศาสตร์ของข้อทดสอบภาษาอังกฤษ เข้ามหาวิทยาลัย ชุด กข. และ กขค. ปี พ.ศ. 2531-2533 เมื่อวิเคราะห์โดย
 1. วิธีการหนดจุดเดลต้า (Delta Plot Method: DPM)
 2. วิธีทดสอบความแตกต่างของสัดส่วนระหว่างกลุ่มด้วยไคสแควร์ (Chi-square Method: CSM) และ
 3. วิธีการตอบสนองของข้อทดสอบชนิด 3 พารามิเตอร์ (Three-Parameter Logistic Method: TPLM)
- ข. เปรียบเทียบจำนวนข้อทดสอบที่มีความลำเอียงต่อเพศและต่อภาคภูมิศาสตร์ เมื่อวิเคราะห์ด้วยทั้ง 3 วิธีดังกล่าวแล้ว

ต่อไปนี้คือรายละเอียดของผลการวิจัยทั้ง 2 ส่วนที่สำคัญดังกล่าวแล้ว

- ก. ความลำเอียงต่อเพศและต่อภาคภูมิศาสตร์ของข้อทดสอบภาษาอังกฤษ เข้ามหาวิทยาลัย ชุด กข. และ กขค. ปี พ.ศ. 2531-2533
 1. วิเคราะห์ด้วยวิธีการหนดจุดเดลต้า (Delta-Plot Method)

จากการคำนวณหาระยะห่างตั้งฉากของคู่อันดับเดลต้าไปยังแกนหลัก (d) ของข้อทดสอบภาษาอังกฤษชุด กข. และ กขค. ปี พ.ศ. 2531-2533 เมื่อทำการทดสอบโดยกลุ่มผู้สอบเพศชายและเพศหญิง และผู้สอบจากภาคภูมิศาสตร์ต่าง ๆ ปรากฏผลตามรายละเอียดดังในตารางที่ 1.1-1.12 ในภาคผนวก ก. และสามารถสรุปผลการค้นพบได้ดังตารางต่อไปนี้

 - 1.1 ความลำเอียงต่อเพศของข้อทดสอบภาษาอังกฤษ ชุด กข. ปี พ.ศ. 2531-2533

จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4: จำนวนข้อทดสอบที่มีความลำเอียงต่อเพศของแบบทดสอบภาษาอังกฤษ ชุด กข. ปี พ.ศ. 2531-2533 เมื่อวิเคราะห์ด้วย Delta-Plot Method

ปี พ.ศ.	แบบทดสอบชุด	n ลำเอียง*	ลักษณะ** การลำเอียง	ข้อที่ลำเอียงต่อเพศ
2531	กข.	3	M = 1 F = 2	28 59, 91
2532	กข.	13	M = 7 F = 6	23, 24, 39, 51, 55, 97, 99 34, 63, 65, 76, 77, 88
2533	กข.	4	M = 1 F = 3	85 14, 69, 100
$\bar{X}, \approx \bar{X}$		6.67	M = 3 F = 4	

* $r_{xy} = 0.9742 - 0.9831$ และ เกณฑ์ $(\pm 1.964 S_D) = \pm 0.2554$ ถึง ± 0.3152

** M = เพศชาย, F = เพศหญิง

จากตารางที่ 4 ปรากฏว่าเมื่อวิเคราะห์หาความลำเอียงของข้อทดสอบด้วยวิธีกำหนดจุดเดลต้าแล้ว พบว่า โดยทั่ว ๆ ไปแล้ว แบบทดสอบภาษาอังกฤษ ชุด กข. ปี พ.ศ. 2531-2533 แต่ละชุดมีข้อทดสอบที่มีความลำเอียงต่อเพศน้อยข้อ กล่าวคือมีชุดละ 3-13 ข้อเท่านั้น แบบทดสอบที่มีข้อทดสอบที่มีความลำเอียงต่อเพศน้อยที่สุด คือแบบทดสอบภาษาอังกฤษชุด กข. ปี พ.ศ. 2531 ซึ่งมีข้อทดสอบที่ลำเอียงต่อเพศเพียง 3 ข้อเท่านั้น ส่วนแบบทดสอบที่มีความลำเอียงต่อเพศมากที่สุด คือแบบทดสอบภาษาอังกฤษ ชุด กข. ปี พ.ศ. 2532 ซึ่งพบว่ามีข้อทดสอบที่มีความลำเอียงต่อเพศ 13 ข้อ และโดยเฉลี่ยแล้วแบบทดสอบแต่ละชุดมีข้อสอบที่มีความลำเอียงต่อเพศประมาณ 6-7 ข้อ

อึ้ง เมื่อจำแนกลักษณะของความล้มอึ้งต่อเพศแล้ว พบว่าโดยทั่ว ๆ ไปแบบทดสอบทั้ง 3 ฉบับ มีข้อทดสอบที่มีความล้มอึ้งต่อเพศชายและเพศหญิงในจำนวนที่ใกล้เคียงกันมาก คือ ประมาณ 3-4 ข้อต่อฉบับ

นอกจากนี้ยังมีข้อนำสังเกตว่าค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าเดลต้า (Δ) ซึ่งเป็นค่าระดับความยากง่ายของข้อทดสอบแต่ละข้อของผู้สอบเพศชายและเพศหญิงมีค่าสูงมาก กล่าวคือมีค่าระหว่าง 0.9742-0.9831 จึงทำให้เกณฑ์ที่บ่งบอกความล้มอึ้งเมื่อ $\pm 1.964 S_d$ ($p = 0.05$) มีค่าต่ำ

1.2 ความล้มอึ้งต่อเพศของข้อทดสอบภาษาอังกฤษชุด กขค. ปี พ.ศ. 2531-2533

ตารางที่ 5: จำนวนข้อทดสอบที่มีความล้มอึ้งต่อเพศของแบบทดสอบภาษาอังกฤษชุด กขค. ปี พ.ศ. 2531-2533 เมื่อวิเคราะห์ด้วย Delta Plot Method

ปี พ.ศ.	แบบทดสอบชุด	n ล้มอึ้ง*	ลักษณะการล้มอึ้ง	ข้อที่ล้มอึ้งต่อเพศ
2531	กขค.	6	M = 3	16, 91, 96
			F = 3	31, 76, 90
2532	กขค.	4	M = 2	4, 8
			F = 2	30, 75
2533	กขค.	2	M = 1	16
			F = 1	23
$\bar{X}_m \approx \bar{X}_f$		4.00	M = 2 F = 2	

* $r_{xy} = 0.9610 - 0.9719$ และ เกณฑ์ ($\pm 1.964 S_d$) = ± 0.3288 ถึง ± 0.3878

จากตารางที่ 5 ปรากฏว่าเมื่อวิเคราะห์หาความลำเอียงของข้อทดสอบด้วยวิธีกำหนดจุดเดลต้าแล้วพบว่า โดยทั่ว ๆ ไปแล้ว แบบทดสอบภาษาอังกฤษ ชุด กขค. ปี พ.ศ. 2531-2533 แต่ละฉบับมีข้อทดสอบที่มีความลำเอียงต่อเพศน้อยข้อ กล่าวคือมีฉบับละ 2-6 ข้อเท่านั้น หรือโดยเฉลี่ยแล้วประมาณ 4 ข้อต่อฉบับ แบบทดสอบที่มีความลำเอียงต่อเพศน้อยที่สุด ได้แก่ แบบทดสอบภาษาอังกฤษ ชุด กขค. ปี พ.ศ. 2533 ซึ่งมีข้อทดสอบที่ลำเอียงเพียง 2 ข้อเท่านั้น ส่วนแบบทดสอบที่มีความลำเอียงมากที่สุด ได้แก่แบบทดสอบภาษาอังกฤษชุด กขค. ปี พ.ศ. 2531 ซึ่งมี 6 ข้อ

อนึ่ง ปรากฏว่าแบบทดสอบมีความลำเอียงต่อเพศชายและเพศหญิง โดยเฉลี่ยแล้วเท่ากัน (1:1) และโดยเฉลี่ยแล้วแบบทดสอบภาษาอังกฤษ ชุด กขค. ปี พ.ศ. 2531-2533 มีข้อสอบที่มีความลำเอียงต่อเพศชายหรือเพศหญิงประมาณฉบับละ 2 ข้อ

นอกจากนี้ยังพบว่าค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าเดลต้า (Δ) ของผู้สอบเพศชายและเพศหญิงมีค่าสูงมาก คือ 0.9610-0.9719 จึงทำให้เกณฑ์การตัดสินความลำเอียงมีค่าต่ำ และโดยทั่ว ๆ ไปแล้ว มีค่าสูงกว่าเกณฑ์ในการตัดสินความลำเอียงต่อเพศของแบบทดสอบชุด กข.

1.3 ความลำเอียงต่อภาคภูมิศาสตร์ของข้อทดสอบภาษาอังกฤษชุด กข.
ปี พ.ศ. 2531-2533

ตารางที่ 6 จำนวนข้อทดสอบที่มีความลำเอียงต่อผู้สอบจากภาคภูมิศาสตร์ต่าง ๆ ของแบบทดสอบภาษาอังกฤษ ชุด กข. ปี พ.ศ. 2531-2533 เมื่อวิเคราะห์ด้วย Delta Plot Method

ปี พ.ศ.	แบบทดสอบชุด	n ลำเอียง	ลักษณะการลำเอียง	ข้อที่ลำเอียงต่อแต่ละภาค	ข้อที่ลำเอียงต่อภาค
2531	กข.	6	C = 3	14, 19, 26	14, 26, 50, 67 (4)
		(C vs E)	E = 3	50, 53, 85	
		9	C = 0	-	
		(C vs NE)	NE = 9	3, 7, 10, 13 15, 24, 29, 55, 89	

(ตารางที่ 6 ต่อ)

ปี พ.ศ.	แบบ ทดสอบชุด	n ลำเอียง	ลักษณะ การลำเอียง	ข้อที่ลำเอียงต่อ แต่ละภาค	ข้อที่ลำเอียงต่อภาค
2532	กข	7 (C vs N)	C = 4 N = 3	17, 19, 67, 80, 50, 56, 96	
		8 (C vs S)	C = 4 S = 4	9, 14, 26, 67 33, 49, 50, 86	
		11 (C vs E)	C = 5 E = 6	12, 19, 35, 76, 89 3, 7, 37, 38, 40, 42	3, 7, 19, 35, 37, 38, 40,
		10 (C vs NE)	C = 4 NE = 6	19, 29, 75, 76 5, 7, 10, 43, 64, 69	43, 64, 69, 76 (11)
		13 (C vs N)	C = 6 N = 7	2, 6, 52, 57, 69, 76 34, 38, 38, 40, 45, 64, 82	
		13 (C vs S)	C = 6 S = 7	19, 33, 35, 44, 69, 76 3, 7, 37, 38, 40, 43, 64	



ตารางที่ 6 (ต่อ)

ปี พ.ศ.	แบบ ทดสอบชุด	n ลำเอียง	ลักษณะ การลำเอียง	ข้อที่ลำเอียงต่อ แต่ละภาค	ข้อที่ลำเอียงต่อภาค
2533	กข.	5	C = 1	92	12, 42, 73 (3)
		(C vs E)	E = 4	17, 42, 76, 84	
		6	C = 3	12, 31, 96	
		(C vs NE)	NE = 3	11, 73, 75	
2533	กข.	8	C = 4	12, 21, 64, 89	
		(C vs N)	N = 4	20, 33, 65, 73	
2533	กข.	6	C = 1	31	
		(C vs S)	S = 5	11, 22, 42, 65, 75	
$\bar{X}, \approx \bar{X}$		8.50	C = 3 E = 4 NE = 6 N = 5 S = 5		6.00

* C = ภาคกลาง E = ภาคตะวันออก NE = ภาคตะวันออกเฉียงเหนือ
N = ภาคเหนือ S = ภาคใต้

จากตารางที่ 6 ปรากฏว่าโดยเฉลี่ยแล้วแบบทดสอบภาษาอังกฤษ ชุด กข. ปี พ.ศ. 2531-2533 มีข้อทดสอบที่ลำเอียงต่อแต่ละภาคประมาณ 8-9 ข้อ ข้อทดสอบเหล่านี้มีความลำเอียงต่อภาค (กล่าวคือลำเอียงต่อผู้สอบมากกว่า 1 ภาคภูมิศาสตร์ขึ้นไป) ประมาณ 6 ข้อ เมื่อพิจารณาค่าเฉลี่ยโดยประมาณ ของจำนวนข้อทดสอบที่ลำเอียงต่อผู้สอบในภาคต่าง ๆ แล้ว ปรากฏว่าแบบทดสอบภาษาอังกฤษแต่ละฉบับ มีข้อทดสอบที่ลำเอียงต่อผู้สอบจากภาคกลางน้อยที่สุด คือ 3 ข้อ รองลงมาได้แก่ลำเอียงต่อผู้สอบภาคตะวันออก คือ 4 ข้อ อันดับที่ 3 ได้แก่ข้อสอบที่ลำเอียงต่อผู้สอบจากภาคเหนือและภาคใต้ซึ่งมีฉบับละประมาณ 5 ข้อ และข้อทดสอบที่มีความลำเอียงต่อผู้สอบจากภาคตะวันออกเฉียงเหนือมากที่สุด กล่าวคือประมาณฉบับละ 6 ข้อ

อนึ่ง เมื่อพิจารณาจำนวนข้อทดสอบที่ลำเอียงต่อผู้สอบจากภาคต่าง ๆ ในแต่ละปีการศึกษา และแต่ละฉบับแล้ว ปรากฏว่าแบบทดสอบภาษาอังกฤษชุด กข. ปี พ.ศ. 2533 มีข้อทดสอบที่มีความลำเอียงต่อภาคต่าง ๆ จำนวนน้อยที่สุด คือระหว่าง 5-8 ข้อ แต่แบบทดสอบภาษาอังกฤษ ชุด กข. ในปี พ.ศ. 2532 มีข้อทดสอบที่มีความลำเอียงต่อภาคต่าง ๆ จำนวนมากที่สุด คือระหว่าง 10-13 ข้อ และเป็นที่น่าสนใจเกี่ยวกับแบบทดสอบภาษาอังกฤษชุด กข. ปี พ.ศ. 2531 มีข้อทดสอบที่มีความลำเอียงต่อผู้สอบจากภาคตะวันออกเฉียงเหนือมากที่สุด ในจำนวนแบบทดสอบทั้งหมด 3 ฉบับ ใน 3 ปี กล่าวคือ มี 9 ข้อ และยังพบว่าแบบทดสอบภาษาอังกฤษชุด กข. ปี 2532 มีความลำเอียงต่อภาคมากที่สุดด้วย คือมีทั้งหมด 11 ข้อ ในขณะที่แบบทดสอบอีก 2 ฉบับ มีความลำเอียงต่อภาคเพียง 3-4 ข้อเท่านั้น

1.4 ความลำเอียงต่อภาคภูมิศาสตร์ของข้อทดสอบภาษาอังกฤษ ชุด กขค.
ปี พ.ศ. 2531-2533

สถาบันวิจัยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 7 จำนวนข้อทดสอบที่มีความลำเอียงต่อผู้สอบจากภาคภูมิศาสตร์ต่าง ๆ ของ
แบบทดสอบภาษาอังกฤษ ชุด กขค. ปี พ.ศ. 2531-2533 เมื่อวิเคราะห์ด้วย
Delta-Plot Method

ปี พ.ศ.	แบบ ทดสอบชุด	n ลำเอียง	ลักษณะ การลำเอียง	ข้อที่ลำเอียงต่อ แต่ละภาค	ข้อที่ลำเอียงต่อภาค
2531	กขค.	8	C = 6	10, 14, 30,	14, 30, 44, 48, 51
		(C vs E)	E = 2	44, 51, 62	
		4	C = 1	24, 48	(5)
		(C vs NE)	NE = 3	14	
		9	C = 5	12, 30, 40	
		(C vs N)	N = 4	17, 30, 35, 44, 51	
		2	C = 1	13, 48, 63, 74	
(C vs S)	S = 1	44			
2532	กขค.	4	C = 2	83, 94	19, 24, 27 (3)
		(C vs E)	E = 2	27, 40	
		4	C = 0	-	
		(C vs NE)	NE = 4	11, 19, 24, 27	
		6	C = 3	29, 57, 87	
		(C vs N)	N = 3	19, 24, 27	
2	C = 0	-			
(C vs S)	S = 2	15, 19			

ตารางที่ 7 (ต่อ)

ปี พ.ศ.	แบบ ทดสอบชุด	n ลำเอียง	ลักษณะ การลำเอียง	ข้อที่ลำเอียงต่อ แต่ละภาค	ข้อที่ลำเอียงต่อภาค
2533	กขค.	8 (C vs E)	C = 2 E = 6	72, 98 15, 63, 67, 83, 87, 90	15, 53, 67, 71, 72, 87, 90, 98 (8)
		22 (C vs NE)	C = 6 NE = 16	51, 52, 64, 70, 81, 91 9, 50, 53, 57, 61, 62, 67, 69, 71, 75, 76, 93, 94, 95, 96, 99	
		8 (C vs N)	C = 1 N = 7	71 15, 53, 67, 72, 73, 87, 98	
		9 (C vs S)	C = 1 S = 8	72 15, 58, 67, 71, 78, 87, 90, 98	
$\bar{X}, \approx \bar{X}$		7.16	C = 2 E = 3 NE = 8 N = 5 S = 4		5.33

จากตารางที่ 7 ปรากฏว่าโดยเฉลี่ยแล้วแบบทดสอบภาษาอังกฤษ ชุด กขค. ปี พ.ศ. 2531-2533 มีข้อทดสอบที่มีความลำเอียงต่อแต่ละภาคประมาณ 7 ข้อ ข้อทดสอบเหล่านี้มีความลำเอียงต่อภาค (ลำเอียงต่อผู้สอบมากกว่า 1 ภาค) ประมาณฉบับละ 5 ข้อ เมื่อพิจารณาค่าเฉลี่ยโดยประมาณของจำนวนข้อทดสอบที่ลำเอียงต่อแต่ละภาคแล้ว ปรากฏว่าแบบทดสอบแต่ละฉบับมีข้อทดสอบที่ลำเอียงต่อผู้สอบจากภาคกลางน้อยที่สุด คือ 2 ข้อ รองลงมาได้แก่ลำเอียงต่อผู้สอบจากภาคตะวันออก คือ 2 ข้อ อันดับที่ 2 ได้แก่ ลำเอียงต่อผู้สอบจากภาคใต้ คือ 4 ข้อ และข้อทดสอบที่มีความลำเอียงต่อผู้สอบจากภาคตะวันออกเฉียงเหนือ จำนวนมากที่สุด คือประมาณ 8 ข้อต่อฉบับ

อนึ่ง เมื่อพิจารณาจำนวนข้อทดสอบที่ลำเอียงต่อผู้สอบจากภาคต่าง ๆ ในแต่ละปีการศึกษา และแต่ละฉบับแล้ว ปรากฏว่าแบบทดสอบภาษาอังกฤษ ชุด กขค. ปี พ.ศ. 2532 มีข้อทดสอบที่มีความลำเอียงต่อแต่ละภาคน้อยที่สุด คือประมาณ 2-6 ข้อเท่านั้น แต่แบบทดสอบภาษาอังกฤษ ชุด กขค. ปี พ.ศ. 2533 มีข้อทดสอบที่มีความลำเอียงต่อแต่ละภาคมากที่สุด คือประมาณ 8-22 ข้อ และเป็นที่น่าสังเกตว่าแบบทดสอบดังกล่าวนี้มีความลำเอียงต่อผู้สอบจากภาคตะวันออกเฉียงเหนือมากที่สุด คือ 16 ข้อ และเป็นแบบทดสอบที่มีความลำเอียงต่อภาคมากที่สุด คือ 8 ข้อ ขณะที่แบบทดสอบชุด กขค. ของปี พ.ศ. 2531-2533 มีข้อทดสอบที่ลำเอียงต่อภาคเพียง 3-5 ข้อเท่านั้น

ดังนั้น จากตารางที่ 6-7 อาจกล่าวได้ว่าการวิเคราะห์หาความลำเอียงของข้อทดสอบของแบบทดสอบภาษาอังกฤษชุด กข. และ กขค. ปี พ.ศ. 2531-2533 พบว่าแบบทดสอบภาษาอังกฤษ ชุด กข. ปี พ.ศ. 2532 มีข้อทดสอบที่ลำเอียงต่อภาคมากที่สุดคือ 11 ข้อ รองลงมาได้แก่แบบทดสอบภาษาอังกฤษ ชุด กขค. ปี พ.ศ. 2533 ซึ่งมีข้อทดสอบที่ลำเอียงต่อภาค 8 ข้อ และแบบทดสอบภาษาอังกฤษชุด กข. ปี พ.ศ. 2533 และชุด กขค. ปี พ.ศ. 2532 เป็นแบบทดสอบที่พบที่มีความลำเอียงน้อยที่สุดคือ 2 ข้อต่อฉบับเท่านั้น และอาจกล่าวด้วยได้ว่าแบบทดสอบโดยมากมีความลำเอียงต่อผู้สอบจากภาคตะวันออกเฉียงเหนือมากที่สุด รองลงมาได้แก่ผู้สอบจากภาคเหนือ และข้อทดสอบที่มีความลำเอียงต่อผู้สอบจากภาคกลางน้อยที่สุด

2. วิเคราะห์ด้วยวิธีทดสอบความแตกต่างของสัดส่วนระหว่างกลุ่มด้วย ไคสแควร์ (Chi-square method)

จากการคำนวณหาความแตกต่างอย่างมีนัยสำคัญของสัดส่วนระหว่างกลุ่มผู้สอบเมื่อจำแนกระดับความสามารถออกเป็น 3 พวกคือ ระดับต่ำ ระดับปานกลาง และระดับสูง และจำแนกตามเพศและภาคภูมิศาสตร์ของผู้สอบ แล้วจึงทดสอบความแตกต่างดังกล่าวด้วยไคสแควร์ (Chi-square test) ที่ระดับความมีนัยสำคัญที่ $\alpha = 0.05$ ปรากฏผลตามรายละเอียดดังในตารางที่ 2.1 - 2.12.4 ในภาคผนวก ข. และสามารถสรุปผลการค้นพบได้ดังตารางต่อไปนี้

2.1 ความลำเอียงต่อเพศของข้อทดสอบภาษาอังกฤษ ชุด กข. ปี พ.ศ.
2531-2533

ตารางที่ 8: จำนวนข้อทดสอบที่มีความลำเอียงต่อเพศของแบบทดสอบภาษาอังกฤษ ชุด กข.
ปี พ.ศ. 2531-2533 เมื่อวิเคราะห์ด้วย Chi-square Method

ปี พ.ศ.	แบบทดสอบชุด	n ลำเอียง*	ลักษณะ** การลำเอียง	ข้อที่ลำเอียงต่อเพศ
2531	กข.	16	M = 9 F = 6 B = 1	4, 13, 17, 33, 43, 60, 61, 85, 92 25, 63, 73, 94, 95, 99 1
2532	กข.	28	M = 13 F = 14 B = 1	16, 19, 28, 34, 40, 43, 45, 51, 69, 71, 82, 87, 96 2, 4, 7, 14, 39, 41, 49, 56, 57, 63, 76, 81, 97, 98 59
2533	กข.	9	M = 5 F = 4 B = 0	1, 34, 37, 59, 81 17, 33, 43, 100 -
$\bar{X}, \approx \bar{X}$		17.66	M = 9 F = 8 B = 1	

* เกณฑ์ $X^2_{1, .017} = 5.7242, \alpha = 0.05$

** B = ทั้งเพศชายและเพศหญิง

จากตารางที่ 8 ปรากฏว่าแบบทดสอบภาษาอังกฤษ ชุด กข. ปี พ.ศ. 2533 เป็นแบบทดสอบที่มีข้อทดสอบที่ลำเอียงต่อเพศน้อยที่สุด คือ 9 ข้อ แต่แบบทดสอบ ชุด กข. ปี พ.ศ. 2532 มีข้อทดสอบที่ลำเอียงต่อเพศมากที่สุด คือ 28 ข้อ โดยเฉลี่ยแล้วแบบทดสอบชุด กข. ปี พ.ศ. 2531-2533 มีข้อทดสอบที่ลำเอียงต่อเพศ ประมาณ 17-18 ข้อต่อฉบับ ข้อทดสอบเหล่านี้ลำเอียงต่อเพศชายและเพศหญิงจำนวนใกล้เคียงกัน คือลำเอียงต่อเพศชาย ประมาณ 9 ข้อ และลำเอียงต่อเพศหญิงประมาณ 8 ข้อ อีก 1 ข้อ มีความลำเอียงต่อเพศชายและเพศหญิง กล่าวคือในบางระดับความสามารถ ผู้สอบเพศหญิงหรือเพศชายอาจได้เปรียบกว่าผู้สอบอีกเพศหนึ่ง แต่ในบางระดับความสามารถก็อาจเสียเปรียบได้ด้วย

2.2 ความลำเอียงต่อเพศของข้อทดสอบภาษาอังกฤษชุด กขด. ปี พ.ศ. 2531-2533

ตารางที่ 9: จำนวนข้อทดสอบที่มีความลำเอียงต่อเพศของแบบทดสอบภาษาอังกฤษ ชุด กขด. ปี พ.ศ. 2531-2533 เมื่อวิเคราะห์ด้วย Chi-square Method

ปี พ.ศ.	แบบทดสอบชุด	n ลำเอียง*	ลักษณะ** การลำเอียง	ข้อที่ลำเอียงต่อเพศ
2531	กขด.	22	M = 15 F = 7 B = 0	16, 17, 20, 24, 25, 28, 49, 52, 55, 56, 62, 79, 91, 96, 99 31, 33, 50, 67, 69, 76, 90 -
2532	กขด.	17	M = 12 F = 5 B = 0	7, 10, 11, 13, 14, 19, 28, 34, 48, 55, 65, 67 30, 47, 52, 78, 84 -

ตารางที่ 9 (ต่อ)

ปี พ.ศ.	แบบทดสอบชุด	n ลำเอียง*	ลักษณะ** การลำเอียง	ข้อที่ลำเอียงต่อเพศ
2533	กขค.	14	M = 9 F = 5 B = 0	12, 15, 16, 30, 36, 67, 79, 97, 98 3, 7, 20, 23, 48 -
$\bar{X}, \approx \bar{X}$		17.67	M = 12 F = 6 B = 0	

จากตารางที่ 9 ปรากฏว่าแบบทดสอบภาษาอังกฤษชุด กขค. ปี พ.ศ. 2533 เป็นแบบทดสอบที่มีข้อทดสอบที่ลำเอียงต่อเพศน้อยที่สุดคือ 14 ข้อ แต่แบบทดสอบภาษาอังกฤษชุด กขค. ปี พ.ศ. 2531 มีข้อทดสอบที่ลำเอียงต่อเพศมากที่สุดคือ 22 ข้อ โดยเฉลี่ยแล้วแบบทดสอบภาษาอังกฤษชุด กขค. ปี พ.ศ. 2351-2533 มีข้อสอบที่ลำเอียงต่อเพศประมาณ 17-18 ข้อต่อฉบับ แต่ข้อทดสอบเหล่านี้มีความลำเอียงต่อผู้สอบเพศชายมากกว่าต่อเพศหญิงในอัตราส่วน 2:1 (12:6) กล่าวคือลำเอียงต่อเพศชาย 12 ข้อ และลำเอียงต่อเพศหญิง 6 ข้อ การวิเคราะห์ข้อทดสอบโดยใช้ Chi-square Method ครั้งนี้ ไม่พบว่าข้อทดสอบใดที่มีความลำเอียงต่อเพศชายและเพศหญิงในต่างระดับความสามารถ

2.3 ความลำเอียงต่อภาคภูมิศาสตร์ของข้อทดสอบภาษาอังกฤษ ชุด กข. ปี พ.ศ. 2531-2533

ตารางที่ 10: จำนวนข้อทดสอบที่มีความลำเอียงต่อผู้สอบจากภาคภูมิศาสตร์ต่าง ๆ ของแบบทดสอบภาษาอังกฤษ ชุด กข. ปี พ.ศ. 2531-2533 เมื่อวิเคราะห์ด้วย Chi-square Method

ปี พ.ศ.	แบบ ทดสอบชุด	n ลำเอียง	ลักษณะ* การลำเอียง	ข้อที่ลำเอียงต่อ แต่ละภาค	ข้อที่ลำเอียงต่อภาค
2531	กข.	40 (C vs E)	C = 7 E = 32	14, 16, 29, 57, 61, 87, 93 2-4, 13, 15, 17, 18, 28, 30-33, 38, 39, 42, 48, 50, 51-53, 59, 63, 66, 68, 74, 80, 85, 86, 88- 90, 94	1, 2, 4, 7, 13-17, 21, 22, 26, 29, 30, 32, 33, 36, 38, 39, 42, 43, 48, 50, 52, 57, 59, 61, 63, 68, 80, 85-90, 94, 96 (37)
		41 (C vs NE)	B = 1 C = 18 NE = 21	26 1, 2, 6, 12, 14, 16, 20-22 26, 30, 34, 36, 44, 58, 61, 95, 99 4, 10, 15, 29, 33, 38, 39, 42, 48, 50, 52, 57, 59, 62, 68, 80, 85, 86, 89, 90, 96	

(ตารางที่ 10) ต่อ

ปี พ.ศ.	แบบ ทดสอบชุด	n ลำเอียง	ลักษณะ* การลำเอียง	ข้อที่ลำเอียงต่อ แต่ละภาค	ข้อที่ลำเอียงต่อภาค
		10 (C vs N)	B = 2 C = 4 N = 6	17, 63 7, 14, 19, 43 29, 42, 48, 50, 90, 98	
		39 (C vs S)	B = 0 C = 13 S = 24	- 1, 11, 13, 14, 21, 22, 26, 36, 49, 57, 67, 76, 87, 4, 5, 16, 17, 29, 32, 33, 38, 41-43, 46, 48, 50, 59, 78, 83- 86, 88, 89, 94, 96	
2532	กข.	55 (C vs E)	B = 2 C = 7 E = 43	2, 7 18, 22, 39, 45, 57, 65, 66 2, 5-8, 10- 14, 17, 25, 26, 28, 31- 33, 36, 37, 40, 41, 43, 44, 46, 50, 51, 53, 55,	2, 5, 6, 8, 10, 11, 13, 14, 16- 18, 21-23, 25, 26, 28, 30, 32, 36-43, 45, 46, 50-53, 57, 60, 61, 64, 66, 67, 70, 72, 76, 77 80-82, 87, 92, 95, 96, 98, 99 (52)

(ตารางที่ 10) ต่อ

ปี พ.ศ.	แบบ ทดสอบชุด	n ลำเอียง	ลักษณะ* การลำเอียง	ข้อที่ลำเอียงต่อ แต่ละภาค	ข้อที่ลำเอียงต่อภาค
		55 (C cs NE)	B = 5 C = 13 NE = 35	57, 59-61, 64, 70, 72, 76, 80-82, 87, 91, 93, 94, 98 30, 38, 42, 52, 96 9, 18, 20, 29, 34, 39, 45, 66- 68, 77, 92, 99 2, 4-6, 8, 11, 14, 16, 17, 19, 23-26, 28, 32, 36, 37, 40, 42, 43, 46, 50, 51, 53, 60, 64, 72, 76, 81, 82, 87, 95, 98	
		11 (C vs N)	C = 3 N = 8	B = 7 C = 3 N = 8	10, 13, 21, 22, 41, 52, 70 66, 81, 99 11, 21, 40, 51, 69, 70, 80, 82
		48 (C vs S)	B = 0 C = 13	B = 0 C = 13	- 10, 35, 38, 39, 41, 45, 47, 57, 67, 77, 84, 92, 99

(ตารางที่ 10) ต่อ

ปี พ.ศ.	แบบ ทดสอบชุด	n ลำเอียง	ลักษณะ* การลำเอียง	ข้อที่ลำเอียงต่อ แต่ละภาค	ข้อที่ลำเอียงต่อภาค
2533	กข.	42 (C vs E)	S = 34 B = 1 C = 9	2, 3, 5, 6, 8, 11, 14, 16, 17, 21, 23, 26, 28, 30, 32, 37, 40, 42, 43, 51-53, 60, 61, 70, 73, 76, 80, 82, 85, 87, 95, 96, 98 13 20, 40, 64, 67, 69, 88, 92, 94, 95	2-4, 7, 10, 11, 13, 14, 16, 17, 19, 20, 22, 24, 26, 28-30, 32-34, 37, 38, 40, 42, 43, 45, 51, 62, 64-67, 71, 74, 75, 76, 79, 80, 86, 88, 91-93, 95 (46)
		45	B = 2 C = 19	5, 13 2, 4, 5, 7, 10, 13, 16, 20, 25, 28, 30, 38, 40, 54, 67, 72, 88, 92, 93	

(ตารางที่ 10) ต่อ

ปี พ.ศ.	แบบ ทดสอบชุด	n ลำเอียง	ลักษณะ* การลำเอียง	ข้อที่ลำเอียงต่อ แต่ละภาค	ข้อที่ลำเอียงต่อภาค
		23 (C vs N)	NE = 20 B = 6 C = 6 N = 5	3, 14, 17, 19, 22, 24, 37, 43, 45, 51, 60, 65, 66, 74, 75, 79, 80, 86, 91, 98 11, 26, 29, 33, 42, 76 8, 21, 58, 64, 88, 89 3, 13, 17, 20, 22, 24, 26, 32- 34, 37, 66, 71, 82, 85	
		36 (C vs S)	B = 2 C = 8 S = 27	1, 30 2, 7, 10, 20, 29, 34, 38, 93 4, 6, 11, 14, 16, 17, 19, 22-24, 26, 30, 32, 33, 42, 45, 51, 62, 65, 66, 71, 74, 76, 80, 86, 91, 95	

(ตารางที่ 10) ต่อ

ปี พ.ศ.	แบบ ทดสอบชุด	n ลำเอียง	ลักษณะ* การลำเอียง	ข้อที่ลำเอียงต่อ แต่ละภาค	ข้อที่ลำเอียงต่อภาค
			B = 1	13	
$\bar{X} \approx \bar{X}$		37.08	C = 10 E = 35 NE = 25 N = 10 S = 28 E = 2		45.00

* B = ทั้ง 2 ภาค เช่น ภาคกลาง (C) และภาคตะวันออกเฉียงเหนือ (E)

จากตารางที่ 10 ปรากฏว่าโดยเฉลี่ยแล้วแบบทดสอบภาษาอังกฤษ ชุด กข. ปี พ.ศ. 2531-2533 มีข้อทดสอบที่มีความลำเอียงต่อแต่ละภาค ประมาณฉบับละ 37 ข้อ และมีข้อทดสอบที่มีความลำเอียงต่อภาคประมาณ 45 ข้อต่อฉบับ ซึ่งนับว่าสูงมาก เมื่อพิจารณาค่าเฉลี่ยโดยประมาณของจำนวนข้อทดสอบที่มีความลำเอียงต่อแต่ละภาคแล้ว ปรากฏว่าแบบทดสอบทั้ง 3 ฉบับดังกล่าวแล้ว มีข้อทดสอบที่ลำเอียงต่อผู้สอบจากภาคตะวันออกเฉียงเหนือมากที่สุด คือประมาณฉบับละ 35 ข้อ รองลงมาได้แก่ข้อทดสอบที่ลำเอียงต่อผู้สอบจากภาคใต้คือประมาณฉบับละ 28 ข้อ และลำเอียงต่อผู้สอบจากภาคกลางน้อยที่สุด คือประมาณ 10 ข้อ ต่อฉบับเท่านั้น รวมทั้งมีข้อทดสอบอีกประมาณ 2 ข้อต่อฉบับที่ลำเอียงต่อผู้สอบจาก 2 ภาคที่มีระดับความสามารถต่างกัน

อนึ่ง เมื่อพิจารณาจำนวนข้อทดสอบที่ลำเอียงต่อผู้สอบจากภาคต่าง ๆ ในแต่ละปีการศึกษา และแต่ละฉบับแล้ว ปรากฏว่าแบบทดสอบภาษาอังกฤษ ชุด กข. ปี พ.ศ. 2531 มีข้อทดสอบที่มีความลำเอียงน้อยที่สุด คือระหว่าง 10-40 ข้อ รวมทั้งมีข้อทดสอบที่ลำเอียงต่อภาคน้อยที่สุด คือ 37 ข้อ แต่แบบทดสอบชุด กข. ปี พ.ศ. 2532 มีข้อทดสอบที่ลำเอียงต่อแต่ละภาคและต่อภาคมากที่สุด คือระหว่าง 11-45 ข้อ และ 52 ข้อตามลำดับ

นอกจากนี้ เป็นที่น่าสังเกตว่าแบบทดสอบทุกฉบับมีความลำเอียงต่อผู้สอบจากภาคตะวันออกเฉียงเหนือมากที่สุด รองลงมาได้แก่ความลำเอียงต่อผู้สอบจากภาคใต้ และลำเอียงต่อผู้สอบจากภาคกลางน้อยที่สุด

2.4 ความลำเอียงต่อภาคภูมิศาสตร์ของข้อทดสอบภาษาอังกฤษ ชุด กขค.
ปี พ.ศ. 2531-2533

ตารางที่ 11: จำนวนข้อทดสอบที่มีความลำเอียงต่อผู้สอบจากภาคภูมิศาสตร์ต่าง ๆ ของ
 แบบทดสอบภาษาอังกฤษ ชุด กขค. ปี พ.ศ. 2531-2533 เมื่อวิเคราะห์ด้วย
 Chi-square Method

ปี พ.ศ.	แบบ ทดสอบชุด	n ลำเอียง	ลักษณะ การลำเอียง	ข้อที่ลำเอียงต่อ แต่ละภาค	ข้อที่ลำเอียงต่อภาค
2531	กขค.	13	C = 0	-	1-3, 12, 13, 20,
		(C vs E)	E = 13	1, 11, 12, 16, 24, 59, 62, 63, 72, 74, 83, 90, 95	24, 26, 38, 44, 48, 49, 58, 59, 62, 63, 68, 69, 72, 76, 78, 83, 88-90, 100
		21	B = 0	-	(26)
		(C vs NE)	C = 1 NE = 18	100 1-4, 20, 21, 24, 26, 39, 48, 57, 62, 63, 69, 78, 83, 89, 90	
		24	B = 2	72, 76	
		(C vs N)	C = 2 N = 22	58, 76 1, 5, 12, 13, 20, 26, 38, 44, 48, 49, 55, 62, 63, 68, 69, 78, 83, 88-90, 98, 99	

(ตารางที่ 11) ต่อ

ปี พ.ศ.	แบบ ทดสอบชุด	n ลำเอียง	ลักษณะ การลำเอียง	ข้อที่ลำเอียงต่อ แต่ละภาค	ข้อที่ลำเอียงต่อภาค
2532	กขค.	32 (C vs S)	B = 0	-	1, 5, 6, 12, 18-24, 27, 28, 33, 37-39, 42, 48, 50, 55, 57, 60-62, 65, 68, 72, 79, 81, 82, 84, 85, 87, 91, 96 (36)
			C = 6	18, 49, 54, 58, 72, 100	
		15 (C vs E)	S = 26	1-3, 8, 13, 24, 29, 33, 38, 41, 44, 46, 48, 59, 62, 64, 67-69, 71, 78, 83, 86, 88-90	
			B = 0	-	
37 (C vs NE)	กขค.	15 (C vs E)	C = 0	-	
			E = 15	2, 5, 21, 27, 33, 37, 42, 46, 57, 61, 65, 82, 83, 90, 96	
		37 (C vs NE)	B = 0	-	
			C = 11	4, 6, 14, 26, 37, 38, 40, 51, 84, 87, 94	
37 (C vs NE)	กขค.	37 (C vs NE)	NE = 25	1, 5, 12, 15, 18-20, 24, 27, 28, 33, 39, 42, 48, 50, 55, 60- 62, 68, 79, 82, 85, 91	

(ตารางที่ 11) ต่อ

ปี พ.ศ.	แบบ ทดสอบชุด	n ลำเอียง	ลักษณะ การลำเอียง	ข้อที่ลำเอียงต่อ แต่ละภาค	ข้อที่ลำเอียงต่อภาค
		28 (C vs N)	B = 1 C = 3 N = 25	22 3, 18, 57 6, 10, 12, 13, 19, 20, 22-24, 27, 29, 31, 48- 50, 55, 65, 68, 72, 79-82, 85, 91	
		33 (C vs S)	B = 0 C = 10 S = 23	- 6, 21, 22, 37, 38, 53, 57, 84, 87, 96 1, 5, 12, 16, 19, 20, 23, 24, 27, 28, 33, 39, 42, 43, 47, 60- 62, 65, 68, 72, 76, 91	
2533	กชค	15 (C vs E)	B = 0 C = 0 E = 15	- - 1, 8, 9, 12, 15, 16, 19, 28, 30, 40, 41, 46, 50, 56, 80	1, 5, 8, 9, 12, 14-17, 21 28, 30, 41, 46, 50, 56, 66, 74 (18)
			B = 0	-	

(ตารางที่ 11) ต่อ

ปี พ.ศ.	แบบ ทดสอบชุด	n ลำเอียง	ลักษณะ การลำเอียง	ข้อที่ลำเอียงต่อ แต่ละภาค	ข้อที่ลำเอียงต่อภาค
		21 (C vs NE)	C = 7 NE = 14	29, 38, 46, 53, 58, 62, 73 8-10, 12, 14- 17, 21, 28, 30, 50, 86, 90	
		20 (C vs N)	B = 0 C = 1 N = 19	- 41 1, 4, 5, 9, 12, 14, 15, 17, 20, 21, 30, 44, 56, 63, 66, 74, 77, 78, 79	
		19 (C vs S)	B = 0 C = 3 S = 16	- 37, 47, 54 5, 8, 9, 12, 14-17, 28, 35, 48-50, 66, 74, 99	
$\bar{X}, \approx \bar{X}$		23.17	C = 4 E = 14 NE = 19 N = 22 S = 22 B = 0		26.67

จากตารางที่ 11 ปรากฏว่าโดยเฉลี่ยแล้วแบบทดสอบภาษาอังกฤษชุด กขค. ปี พ.ศ. 2531-2533 มีข้อทดสอบที่มีความลำเอียงต่อแต่ละภาคประมาณฉบับละ 23 ข้อ และมีข้อทดสอบที่มีความลำเอียงต่อภาคประมาณฉบับละ 26-27 ข้อ มีข้อทดสอบที่ลำเอียงต่อผู้สอบจากภาคเหนือ และภาคใต้มากที่สุด คือประมาณฉบับละ 22 ข้อ รองลงมาได้แก่ลำเอียงต่อผู้สอบจากภาคตะวันออกเฉียงเหนือ คือประมาณ 19 ข้อ และลำเอียงต่อผู้สอบจากภาคกลางน้อยที่สุด คือประมาณฉบับละ 4 ข้อเท่านั้น

อนึ่ง เมื่อพิจารณาจำนวนข้อทดสอบที่ลำเอียงต่อผู้สอบจากภาคต่าง ๆ ในแต่ละปีการศึกษา และแต่ละฉบับแล้ว ปรากฏว่าแบบทดสอบภาษาอังกฤษชุด กขค. ปี พ.ศ. 2533 มีข้อทดสอบที่ลำเอียงน้อยที่สุด คือประมาณ 15-20 ข้อ รวมทั้งมีข้อทดสอบที่ลำเอียงต่อภาคน้อยที่สุดคือ 18 ข้อ แต่แบบทดสอบชุด กขค. ปี พ.ศ. 2532 มีข้อทดสอบที่ลำเอียงต่อแต่ละภาค และต่อภาคมากที่สุด คือระหว่าง 15-37 ข้อ และ 36 ข้อตามลำดับ

นอกจากนี้ เป็นที่น่าสังเกตว่าแบบทดสอบทุกฉบับมีความลำเอียงต่อผู้สอบจากภาคเหนือหรือภาคใต้มากที่สุด และมีความลำเอียงต่อผู้สอบจากภาคกลางน้อยที่สุด

3. วิเคราะห์ด้วยการตอบสนองของข้อทดสอบชนิด 3 พารามิเตอร์ (Three-parameter Logistic Method)

จากตารางคำนวณหาเนื้อที่ (ϕ) ของความแตกต่างระหว่างโค้งลักษณะของข้อทดสอบแต่ละข้อที่มีค่าพารามิเตอร์ 3 ค่า คือ a (ค่าอำนาจจำแนก) b (ค่าความยากง่าย) และ c (ค่าการเดา) แตกต่างกัน และเมื่อผู้สอบมีความสามารถ (θ) ระหว่าง ± 5 และทำการทดสอบโดยกลุ่มผู้สอบเพศชาย และหญิง รวมทั้งจากภาคภูมิศาสตร์ต่าง ๆ ผลปรากฏตามรายละเอียดดังในตารางที่ 3.1 - 3.12 ในภาคผนวก ค. และสามารถสรุปผลการค้นพบได้ดังตารางต่อไปนี้

3.1 ความลำเอียงต่อเพศของข้อทดสอบภาษาอังกฤษ ชุด กข. ปี พ.ศ. 2531-2533

ตารางที่ 12: จำนวนข้อทดสอบที่มีความลำเอียงต่อเพศของแบบทดสอบภาษาอังกฤษ
ชุด กข. ปี พ.ศ. 2531-2533 เมื่อวิเคราะห์ด้วย Three-parameter
Logistic Method

ปี พ.ศ.	แบบทดสอบชุด	n ลำเอียง	ขนาด การลำเอียง	ข้อที่ลำเอียงต่อเพศ
2531	กข.	23	L = 13 M = 10	14, 17, 19, 40, 41, 44, 74, 75, 80, 83, 89, 92, 97 9, 16, 18, 20, 21, 27, 36, 43, 71, 79
2532	กข.	54	L = 31 M = 23	6, 8, 10, 11, 13, 16, 18-20, 24-31, 34, 57, 62, 66, 68, 79-81, 87, 90, 94, 96, 97 9, 12, 21, 33, 38, 40, 43, 49, 55, 56, 60, 63, 67, 70, 71, 77, 78, 83, 84, 89, 92, 95, 98
2533	กข.	8	L = 4 M = 4	26, 59, 60, 69 50, 58, 87, 93
$\bar{X}_s \approx \bar{X}$		28.33	L = 16 M = 12	

* M = ปานกลาง ($\phi > .40-.70$) L = มาก ($\phi > .70$) [ลำเอียงน้อยไม่ได้คิด]

จากตารางที่ 12 ปรากฏว่าแบบทดสอบภาษาอังกฤษ ชุด กข. ปี พ.ศ. 2533 เป็นแบบทดสอบที่มีความลำเอียงต่อเพศน้อยที่สุด คือ 8 ข้อ แต่แบบทดสอบ ชุด กข. ปี พ.ศ. 2532 มีข้อทดสอบที่ลำเอียงต่อเพศมากที่สุดคือ 54 ข้อ โดยเฉลี่ยแล้วแบบทดสอบชุด กข. ปี พ.ศ. 2531-2533 มีข้อทดสอบที่ลำเอียงต่อเพศประมาณ 28 ข้อต่อฉบับ ข้อทดสอบเหล่านี้โดยมากมีความลำเอียงระดับมากจำนวนมากกว่าระดับปานกลาง คือโดยประมาณแล้วมี 16:12 ข้อ หรือ 1.3:1

อนึ่ง โปรดสังเกตว่าจากพื้นที่ของความแตกต่างระหว่างโค้งลักษณะของข้อทดสอบ ไม่สามารถจะบอกได้ว่าข้อทดสอบที่ลำเอียงต่อเพศดังกล่าวแล้วมีความลำเอียงต่อเพศใด เพราะข้อทดสอบข้อเดียวกันอาจมีความลำเอียงต่อเพศใดเพศหนึ่งหรือทั้ง 2 เพศ เมื่อผู้สอบมีระดับความสามารถต่างกันก็ได้ การที่จะทราบลักษณะการลำเอียงของข้อทดสอบแต่ละข้อที่วิเคราะห์โดยวิธี Three-parameter Logistic Method จะต้องนิยามจากกราฟแสดงโค้งลักษณะของข้อทดสอบแต่ละข้อ (เช่นเดียวกับกับกรณีของความลำเอียงต่อภาคภูมิศาสตร์) ซึ่งผู้วิจัยได้แสดงให้เห็นตัวอย่างบางข้อในภาคผนวก ง. แล้ว แต่ไม่ใช่ทั้งหมด เพราะไม่ใช่วัตถุประสงค์หลักของการวิจัยครั้งนี้ เนื่องจากข้อมูลดังกล่าวมีมากเกินไป

3.2 ความลำเอียงต่อเพศของข้อทดสอบภาษาอังกฤษชุด กขค. ปี พ.ศ. 2531-2533

ตารางที่ 13: จำนวนข้อทดสอบที่มีความลำเอียงต่อเพศของแบบทดสอบภาษาอังกฤษ
ชุด กชค. ปี พ.ศ. 2531-2533 เมื่อวิเคราะห์ด้วย Three-parameter
Logistic Method

ปี พ.ศ.	แบบทดสอบชุด	n ลำเอียง	ขนาด การลำเอียง	ข้อที่ลำเอียงต่อเพศ
2531	กชค.	8	L = 4 M = 4	8, 54, 71, 72 25, 45, 76, 90
2532	กชค.	61	L = 39 M = 22	4-13, 15, 19-21, 26-30, 32-35, 40, 46, 50, 53, 55, 65, 66, 82, 84, 88, 89, 91, 94, 97, 98 2, 3, 16, 25, 38, 39, 41, 42, 49, 51, 57, 59, 60, 67-70, 76, 78, 79, 85, 93, 96
2533	กชค.	55	L = 26 M = 29	2, 8, 9, 12, 14, 16, 18, 25, 28, 30, 32, 36, 40, 41, 44, 46, 49, 51, 57, 63, 65, 86, 88, 91, 93, 98 4-7, 11, 13, 17, 19, 20, 24, 34, 37, 38, 45, 48, 50, 58, 59, 61, 64, 75, 77, 80,

(ตารางที่ 13) ต่อ

ปี พ.ศ.	แบบทดสอบชุด	n ลำเอียง	ขนาด การลำเอียง	ข้อที่ลำเอียงต่อเพศ
				85, 90, 92, 94, 96, 100
$\bar{X}, \approx \bar{X}$		41.33	L = 23 M = 18	

จากตารางที่ 13 ปรากฏว่าแบบทดสอบภาษาอังกฤษ ชุด กขค. ปี พ.ศ. 2531 เป็นแบบทดสอบที่มีความลำเอียงต่อเพศน้อยที่สุด คือ 8 ข้อ แต่แบบทดสอบชุด กขค. ปี พ.ศ. 2532 มีข้อทดสอบที่ลำเอียงต่อเพศมากที่สุด คือ 61 ข้อ โดยเฉลี่ยแล้วแบบทดสอบชุด กขค. ปี พ.ศ. 2531-2533 มีข้อทดสอบที่ลำเอียงต่อเพศประมาณ 41 ข้อต่อฉบับ และโดยมากมีความลำเอียงระดับมากกว่าระดับปานกลางคือโดยประมาณแล้ว มี 23:18 ข้อ หรือ 1.3:1

3.3 ความลำเอียงต่อภาคภูมิศาสตร์ของข้อทดสอบภาษาอังกฤษ ชุด กข.
ปี พ.ศ. 2531-2533

สถาบันทฤษฎีบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 14: จำนวนข้อทดสอบที่มีความลำเอียงต่อผู้สอบจากภาคภูมิศาสตร์ต่าง ๆ ของ
แบบทดสอบภาษาอังกฤษ ชุด กข. ปี พ.ศ. 2531-2533 เมื่อวิเคราะห์ด้วย
(Three-parameter Logistic Model)

ปี พ.ศ.	แบบ ทดสอบชุด	n ลำเอียง	ขนาด การลำเอียง	ข้อที่ลำเอียงต่อ แต่ละภาค	ข้อที่ลำเอียงต่อภาค
2531	กข.	28 (C vs E)	L = 16	4, 12, 17, 19,	3, 4, 9, 12, 14,
				33, 34, 44, 50,	16, 17, 19, 22,
				63, 70, 77, 81,	23, 25, 27, 33,
			M = 12	84, 88, 89, 92,	36, 38, 40, 41,
				95	44, 46, 48, 50,
				6, 9, 22, 23,	22, 23, 26, 27,
		32 (C vs NE)	L = 15	38, 40, 60, 79,	33, 36, 38, 40,
				80, 83, 91, 93	41, 44, 46, 48,
				4, 9, 12, 19,	50, 51, 71, 77,
			M = 17	23, 27, 44, 46,	78-83, 88, 89,
				71, 80, 81, 83,	91-93, 95, 97
				89, 92, 97	(36)
24 (C vs N)	L = 13	3, 8, 14, 16,	20, 22, 25, 33,		
		20, 22, 25, 33,	36, 40, 41, 48,		
		36, 40, 41, 48,	50, 77, 82, 93,		
	M = 11	50, 77, 82, 93,	98		
		9, 12, 23, 27,	33, 38, 40, 44,		
		33, 38, 40, 44,	46, 58, 83, 88,		
		46, 58, 83, 88,	89		
		4, 14, 16, 20,	41, 48, 50, 71,		
		41, 48, 50, 71,	78, 91, 97		

(ตารางที่ 14) ต่อ

ปี พ.ศ.	แบบ ทดสอบชุด	n ลำเอียง	ขนาด การลำเอียง	ข้อที่ลำเอียงต่อ แต่ละภาค	ข้อที่ลำเอียงต่อภาค
2532	กข.	34 (C vs S)	L = 21	4, 9, 12, 14, 17, 19, 23-25, 33, 38, 40, 46, 79-81, 83, 87, 89, 92, 97	
			M = 13	3, 11, 16, 36, 55, 67, 71, 74, 76, 77, 82, 85, 95	
		29 (C vs E)	L = 18	1, 10, 17, 26, 28, 42, 44, 45, 57, 71, 74, 83, 87, 89, 90, 95, 96, 99	1, 5, 6, 10, 15- 18, 25, 28-30, 34, 38, 41, 44, 45, 51, 55, 57, 59, 60, 65-68, 71, 74,
		M = 11	2, 6, 16, 18, 25, 51, 60, 64, 65, 81, 100	81, 87, 89, 90, 93, 95-100 (39)	
		53 (C vs NE)	L = 34	6, 8, 10, 13, 16-18, 20, 21, 24, 27, 29-33, 45, 51, 54, 57, 65, 66, 71, 76, 77, 84, 87, 89, 93, 95-97, 99	
		(C vs NE)	M = 19	5, 7, 11, 12, 25, 34, 35, 38, 40, 41, 44, 46, 55, 60, 67, 68,	

(ตารางที่ 14) ต่อ

ปี พ.ศ.	แบบ ทดสอบชุด	n ลำเอียง	ขนาด การลำเอียง	ข้อที่ลำเอียงต่อ แต่ละภาค	ข้อที่ลำเอียงต่อภาค
2533	กข	25 (C vs N)	L = 10	78, 98, 100 5, 6, 10, 45, 59, 71, 89, 94, 95, 98	
		28 (C vs S)	M = 15	4, 16, 17, 25, 28-30, 34, 52, 60, 66, 69, 70, 82, 97	
			L = 13	1, 5, 10, 17, 51, 57, 60, 74, 87, 90, 93, 96, 98	
		45 (C vs E)	L = 19	9, 12, 14, 19, 22, 26, 30, 52, 54, 61, 64, 66, 77, 78, 81, 87, 89, 90, 93	8-10, 12-14, 18- 22, 26, 30, 32, 33, 35, 39, 44, 46-48, 52, 54- 56, 58, 61, 64, 66-69, 72, 73, 77, 78, 81, 87, 89-91, 93, 94, 96, 98, 99
			M = 26	8, 10, 11, 13, 20, 21, 24, 32, 33, 39, 44, 46, 48, 49, 55, 56, 58, 67, 69, 72, 73, 94-96, 98, 99	(46)

(ตารางที่ 14) ต่อ

ปี พ.ศ.	แบบ ทดสอบชุด	n ลำเอียง	ขนาด การลำเอียง	ข้อที่ลำเอียงต่อ แต่ละภาค	ข้อที่ลำเอียงต่อภาค
		55 (C vs NE)	L = 28 M = 27	9, 10, 12, 14, 19, 21, 22, 26- 29, 33, 35, 44, 46, 47, 54, 55, 64, 66, 68, 72, 73, 78, 81, 89, 90, 94 2, 4, 5, 7, 16, 18, 20, 25, 30, 32, 34, 38, 40, 42, 43, 48, 52, 56, 58, 61, 67, 69, 87, 91, 98, 100	
		38 (C vs N)	L = 20	10, 12, 19, 21, 22, 26, 35, 46, 47, 52, 54, 66, 69, 77, 78, 81, 87, 89, 93, 99	
		(C vs N)	M = 18	8, 9, 13, 14, 18, 30, 39, 48, 56, 58, 60, 64, 72, 73, 79, 80, 91, 98	
		33 (C vs S)	L = 15	10, 14, 19, 21, 26, 46, 52, 61, 64, 65, 68, 69, 89, 93, 96	

(ตารางที่ 14) ต่อ

ปี พ.ศ.	แบบ ทดสอบชุด	n ลำเอียง	ขนาด การลำเอียง	ข้อที่ลำเอียงต่อ แต่ละภาค	ข้อที่ลำเอียงต่อภาค
			M = 18	12, 20, 22, 27, 35, 44, 47, 51, 54, 57, 58, 72, 73, 77, 78, 81, 91, 98	
$\bar{X}_1 \approx \bar{X}_2$	C vs E = 25 C vs NE = 47 C vs N = 29 C vs S = 32		L = 19 M = 17		40

จากตารางที่ 14 ปรากฏว่าโดยเฉลี่ยแล้วแบบทดสอบภาษาอังกฤษชุด กข. ปี พ.ศ. 2531-2533 มีข้อทดสอบที่มีความลำเอียงต่อแต่ละภาคประมาณฉบับละ 36 ข้อ และมีความลำเอียงต่อภาคฉบับละประมาณ 40 ข้อ แบบทดสอบมีความลำเอียงต่อผู้สอบจากภาคกลางและภาคตะวันออกเฉียงเหนือมากที่สุด คือประมาณ 47 ข้อ (แต่ไม่ทราบว่าการลำเอียงต่อผู้สอบจากภาคใดมากกว่ากันด้วยเหตุผลดังกล่าวมาแล้ว) รองลงมาได้แก่ ลำเอียงต่อผู้สอบจากภาคกลางและภาคใต้ คือ ประมาณ 32 ข้อ และลำเอียงต่อผู้สอบจากภาคกลางและภาคตะวันออกเฉียงเหนือที่น้อยที่สุดคือ ประมาณ 25 ข้อ

อนึ่ง เมื่อพิจารณาจำนวนข้อทดสอบที่ลำเอียงต่อผู้สอบจากภาคต่าง ๆ ในแต่ละปีการศึกษา และแต่ละฉบับแล้ว ปรากฏว่าแบบทดสอบภาษาอังกฤษ ชุด กข. ปี พ.ศ. 2531 มีข้อทดสอบที่ลำเอียงน้อยที่สุดคือ ประมาณ 24-34 ข้อ และมีข้อทดสอบที่ลำเอียงต่อภาคน้อยที่สุด คือ 36 ข้อ แต่แบบทดสอบชุด กข. ปี พ.ศ. 2533 มีข้อทดสอบที่ลำเอียงต่อแต่ละภาคและต่อภาคมากที่สุด คือ ระหว่าง 33-55 ข้อ และ 46 ข้อ ตามลำดับ

3.4 ความลำเอียงต่อภาคนิศาสตร์ของข้อทดสอบภาษาอังกฤษชุด กขค. ปี พ.ศ. 2531-2533

ตารางที่ 15: จำนวนข้อทดสอบที่มีความลำเอียงต่อผู้สอบจากภาคภูมิศาสตร์ต่าง ๆ ของ
แบบทดสอบภาษาอังกฤษ ชุด กขค. ปี พ.ศ. 2531-2533 เมื่อวิเคราะห์ด้วย
Three-parameter Logistic Model

ปี พ.ศ.	แบบ ทดสอบชุด	n ลำเอียง	ขนาด การลำเอียง	ข้อที่ลำเอียงต่อ แต่ละภาค	ข้อที่ลำเอียงต่อภาค
2531	กข.	60 (C vs E)	L = 29 M = 31	6, 8, 11, 13, 14, 17, 18, 20, 25, 29, 31, 35, 44, 45, 49, 50, 55, 56, 61, 62, 66, 70, 73, 76, 77, 81, 85, 91, 92 1, 2, 5, 7, 9, 10, 15, 21, 22, 34, 36, 39, 46, 51-54, 58, 60, 64, 67, 71, 72, 82, 83, 87, 89, 93, 96, 99	8, 10, 11, 14, 17, 18, 20, 21, 25, 29, 31, 35- 37, 39, 44, 45, 49, 53-56, 58- 62, 66, 67, 70- 74, 76, 77, 79, 81, 82, 84, 85, 87, 89-93, 99, 100 (51)
		42 (C vs NE)	L = 22 M = 20	8, 10, 11, 14, 18, 20, 21, 35, 36, 44, 45, 55, 57, 66, 67, 71, 73, 76, 77, 84, 87, 90 7, 25, 31, 33, 53, 56, 58, 59, 61, 62, 70, 72, 79, 82, 85, 86,	

(ตารางที่ 15) ต่อ

ปี พ.ศ.	แบบ ทดสอบชุด	n ลำเอียง	ขนาด การลำเอียง	ข้อที่ลำเอียงต่อ แต่ละภาค	ข้อที่ลำเอียงต่อภาค
		31 (C vs N)	L = 10 M = 21	91, 93, 99, 100 14, 20, 35, 44, 55, 76, 77, 90, 98, 100 2, 8, 9, 16, 17, 25, 29, 31, 36, 37, 39, 53, 56, 70, 79-81, 85, 93, 97, 99	
		38 (C vs S)	L = 15 M = 23	8, 11, 14, 20, 29, 35, 37, 44, 60, 71, 72, 76, 81, 85, 91 18, 21, 31, 45, 49, 54-56, 58, 60, 62, 66, 70, 73, 74, 77, 82, 86, 90, 92, 94, 99, 100	
2532	กชค	45 (C vs E)	L = 23 M = 22	15, 25, 31, 38, 51, 52, 54, 58, 60, 66, 68, 69, 76, 78, 79, 83, 84, 87, 93, 97- 100 7, 9, 12, 22- 24, 29, 30, 37, 40, 45, 46, 49,	1, 2, 7, 9, 12, 14, 15, 20, 25, 26, 30-32, 35-38, 40, 44, 51, 52, 54, 57, 58, 64, 67, 69, 79, 83, 84, 86, 87, 91, 93-100 (41)



(ตารางที่ 15) ต่อ

ปี พ.ศ.	แบบ ทดสอบชุด	n ลำเอียง	ขนาด การลำเอียง	ข้อที่ลำเอียงต่อ แต่ละภาค	ข้อที่ลำเอียงต่อภาค
		43 (C vs NE)	L = 18 M = 25	55, 56, 59, 61, 64, 86, 90, 92, 96 1, 2, 15, 30, 31, 36, 37, 51, 58, 67, 80, 86- 89, 92, 94, 97 9, 10, 12, 14, 17, 19, 20, 26, 32, 35, 38, 40, 44, 47, 52, 64, 69, 72, 73, 79, 81, 95, 96, 98, 100	
		30 (C vs N)	L = 10 M = 20	8, 15, 31, 32, 54, 57, 79, 92, 94, 95, 6, 7, 9, 11-13, 20, 21, 25, 26, 30, 35, 36, 44, 48, 50, 58, 69, 85, 87	
		29 (C vs N)	L = 18	1, 2, 15, 51, 52, 56, 57, 67, 70, 84, 87, 91, 93, 94, 96, 97, 99, 100	

(ตารางที่ 15) ต่อ

ปี พ.ศ.	แบบ ทดสอบชุด	n ลำเอียง	ขนาด การลำเอียง	ข้อที่ลำเอียงต่อ แต่ละภาค	ข้อที่ลำเอียงต่อภาค
2533	กขค	42 (C vs E)	M = 11	7, 12, 14, 19, 37, 43, 53, 64, 71, 83, 95	4, 8, 15, 21, 23, 26-29, 37, 38, 44, 50, 51, 54, 56, 63, 64, 70, 71, 73, 74, 77, 78, 80, 81, 83, 84, 86, 88, 91, 93, 94, 96, 97, 100 (36)
			L = 17	1, 7, 21, 29, 44, 47, 49, 50, 52, 63, 73, 74, 84, 89, 91, 96, 97	
			M = 25	4, 8, 12, 16, 27, 28, 37, 38, 46, 51, 53, 56, 70, 71, 77, 78, 80, 81, 83, 85, 86, 88, 92-94	
			L = 17	8, 15, 51, 56, 63, 71, 74, 77, 81, 86-89, 91, 93, 96, 97	
			M = 19	3, 10, 14, 23, 25, 27-30, 33, 37, 38, 55, 70, 73, 78, 83, 90, 94	
		19 (C vs N)	L = 12	21, 54, 56, 62, 64, 73, 74, 80, 81, 91, 95, 97	
			M = 7	4, 15, 26, 45, 50, 63, 100	

(ตารางที่ 15) ต่อ

ปี พ.ศ.	แบบ ทดสอบชุด	n ลำเอียง	ขนาด การลำเอียง	ข้อที่ลำเอียงต่อ แต่ละภาค	ข้อที่ลำเอียงต่อภาค
		28 (C vs S)	L = 11 M = 17	7, 23, 44, 64, 71, 74, 84, 89, 91, 96, 97 2, 8, 15, 26, 27, 37, 38, 51, 54, 56, 66, 77, 80, 86, 88, 93, 100	
$\bar{X}_C \approx \bar{X}_S$		C vs E = 49 C vs NE = 40 C vs N = 27 C vs S = 32	L = 17 M = 20		43

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

จากตารางที่ 15 ปรากฏว่าโดยเฉลี่ยแล้วแบบทดสอบภาษาอังกฤษชุด กขค. ปี พ.ศ. 2531-2533 มีข้อทดสอบที่มีความลำเอียงต่อแต่ละภาคประมาณฉบับละ 37 ข้อ และมีความลำเอียงต่อภาคประมาณฉบับละ 43 ข้อ แบบทดสอบที่มีความลำเอียงต่อผู้สอบจากภาคกลางและภาคตะวันออกมากที่สุด คือประมาณ 47 ข้อ รองลงมาได้แก่ ลำเอียงต่อผู้สอบจากภาคกลางและภาคตะวันออกเจียงเหนือ คือ 40 ข้อ และลำเอียงต่อผู้สอบจากภาคกลางและภาคเหนือน้อยที่สุด คือ 27 ข้อ

อนึ่ง เมื่อพิจารณาจำนวนข้อทดสอบที่ลำเอียงต่อผู้สอบจากภาคต่าง ๆ ในแต่ละปีการศึกษา และแต่ละฉบับแล้ว ปรากฏว่าแบบทดสอบภาษาอังกฤษ ชุด กขค. ปี พ.ศ. 2533 มีข้อทดสอบที่ลำเอียงน้อยที่สุดคือประมาณ 11-25 ข้อ และมีข้อทดสอบที่ลำเอียงต่อภาคน้อยที่สุดคือ 36 ข้อ แบบทดสอบชุด กขค. ปี พ.ศ. 2531 มีข้อทดสอบที่ลำเอียงต่อแต่ละภาคมากที่สุด คือประมาณ 10-29 ข้อ มีข้อทดสอบที่มีความลำเอียงต่อภาคมากที่สุด คือ 51 ข้อ

ข. เปรียบเทียบจำนวนข้อทดสอบที่มีความลำเอียงต่อเพศและต่อภาคภูมิศาสตร์เมื่อวิเคราะห์ด้วยทั้ง 3 วิธีดังกล่าวมาแล้ว

1. เปรียบเทียบจำนวนข้อทดสอบที่ลำเอียงต่อเพศของแบบทดสอบภาษาอังกฤษชุด กข. และ กขค. ปี พ.ศ. 2531-2533 เมื่อวิเคราะห์ด้วย 3 วิธี

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 16: เปรียบเทียบจำนวนข้อทดสอบที่มีความลำเอียงต่อเพศของผู้สอบ
เมื่อวิเคราะห์ด้วย 3 วิธี

ปี พ.ศ.	แบบทดสอบ ชุด	n ที่ลำเอียงต่อเพศ			ข้อที่ลำเอียงต่อเพศเกิน 1 วิธี	
		DPM	CSM	TPLM	n	ข้อที่
2531	กข	3	16	23	3	7, 13, 92
2532	กข	13	28	54	22	16, 19, 24, 28, 34, 39, 40, 43, 49, 51, 55-57, 63, 71, 76, 77, 81, 87, 96-98
2533	กข	4	9	8	3	59, 69, 100
\bar{X}		6.67	17.67	28.33	9.33	14.86%
2531	กขค	6	22	8	7	16, 25, 31, 76, 90, 91, 96
2532	กขค	4	17	61	15	4, 7, 8, 10, 11, 13, 19, 28, 30, 34, 55, 65, 67, 78, 84
2533	กขค	2	14	55	9	7, 12, 16, 20, 23, 30, 36, 48, 98
\bar{X}		4.00	17.67	41.33	10.33	16.80%
$\bar{\bar{X}}$		5.33	17.67	34.83	9.83	15.83%
F-test		**	**	**		
r_{xy}		-	0.84*	0.18*		
			-	0.29*		

* $p > 0.05$

** $p < 0.05$

จากตารางที่ 16 ปรากฏว่าโดยเฉลี่ยแล้ว การวิเคราะห์หาความล่าเอียงต่อเพศของข้อทดสอบโดย วิธีการตอบสนองของข้อทดสอบชนิด 3 พารามิเตอร์ (Three-parameter Logistic Method: TPLM) เมื่อวิเคราะห์หาข้อทดสอบที่มีความล่าเอียงสูงกว่าระดับน้อย ($\phi > 0.40$) พบข้อทดสอบที่มีความล่าเอียงจำนวนมากที่สุดคือประมาณ 34-35 ข้อต่อฉบับ รองลงมาได้แก่วิธีทดสอบความแตกต่างของสัดส่วนระหว่างกลุ่มด้วยไคสแควร์ (Chi-square Method: CSM) ซึ่งพบว่าได้ข้อทดสอบที่มีความล่าเอียงแตกต่างจากวิธีนรกอย่างมีนัยสำคัญ คือประมาณ 17-18 ข้อต่อฉบับ การวิเคราะห์หาความล่าเอียงของข้อทดสอบด้วยวิธีกำหนดจุดเดลต้า (Delta Plot Method: DPM) พบข้อทดสอบที่ล่าเอียงจำนวนน้อยที่สุด คือประมาณ 5 ข้อต่อฉบับ และแตกต่างจากอีก 2 วิธีอย่างมีนัยสำคัญ และมีข้อทดสอบประมาณร้อยละ 15-16 ของข้อทดสอบทั้งหมดที่แต่ละวิธีพบว่ามีความล่าเอียงต่อเพศเป็นข้อทดสอบที่พบว่ามีความล่าเอียงต่อเพศ เมื่อทำการวิเคราะห์ด้วยวิธีอื่น ซึ่งอาจได้แก่อีก 1 หรือ 2 วิธี

อนึ่ง ปรากฏว่าความสัมพันธ์ระหว่างจำนวนข้อทดสอบที่ล่าเอียงต่อเพศของแบบทดสอบแต่ละชุดในแต่ละปีของทั้ง 3 วิธี มีระดับค่อนข้างสูงและต่ำมาก แต่ไม่มีนัยสำคัญ ($p = 0.05$)

2. เปรียบเทียบจำนวนข้อทดสอบที่ล่าช้าเชิงต่อภาคภูมิศาสตร์ของแบบทดสอบภาษาอังกฤษ ชุด กข. และ กขค. ปี พ.ศ. 2531-2533 เมื่อวิเคราะห์ด้วย 3 วิธี

ตารางที่ 17: เปรียบเทียบจำนวนข้อทดสอบที่มีความล่าช้าเชิงต่อภาคภูมิศาสตร์ของผู้สอบ เมื่อวิเคราะห์ด้วย 3 วิธี

ปี พ.ศ.	แบบทดสอบชุด	ระหว่าง ภาค	n ที่ล่าช้าเชิงต่อแต่ละภาค			ข้อที่ล่าช้าเชิงต่อภาคเกิน 1 วิธี	
			DPM	CSM	TPLM	n	ข้อที่
2531	กข	C vs E	(4) 6	(37) 40	(36) 28	14	4, 14, 16, 17, 22, 26, 33, 36, 38, 48, 50, 80, 88, 89
		C vs NE	9	41	32		
		C vs N	7	10	24		
		C vs S	8	39	34		
2532	กข	C vs E	(11) 11	(52) 55	(39) 29	28	5, 6, 10, 16-18, 25, 28, 30, 37, 38, 40, 41, 43, 45, 51, 57, 60, 64, 66, 67, 76, 81, 83, 95, 96, 98, 99
		C vs NE	10	55	53		
		C vs N	13	11	25		
		C vs S	13	48	28		
2533	กข	C vs E	(3) 5	(46) 42	(46) 45	18	10, 12-14, 19, 20, 22, 26, 30, 32, 33, 42, 64, 66, 67, 73, 91, 93
		C vs NE	6	45	55		
		C vs N	8	23	38		
		C vs S	6	36	33		
\bar{X}			(6) 8.56	(45) 37.08	(40) 35.33	20.00	21.53%

(ตารางที่ 17) ต่อ

ปี พ.ศ.	แบบทดสอบชุด	ระหว่าง ภาค	n ที่ล่าเอียงต่อแต่ละภาค			ข้อที่ล่าเอียงต่อภาคเกิน 1 วิธี	
			DPM	CSM	TPLM	n	ข้อที่
2531	กชค	C vs E	(5)	(26)	(51)	14	2, 14, 20, 44, 48, 49, 58, 59, 62, 72, 76, 89, 90, 100
			8	13	60		
			4	21	42		
			9	24	31		
2532	กชค	C vs NE	2	32	38	15	1, 12, 18-20, 24, 27, 37, 38, 57, 79, 84, 87, 91, 96
			(3)	(36)	(41)		
			4	15	45		
			4	37	43		
2533	กชค	C vs N	6	28	30	8	8, 15, 21, 28, 50, 56, 71, 74
			2	33	29		
			(8)	(18)	(36)		
			8	15	42		
X̄		C vs S	22	21	36	12.33	16.24%
			8	20	19		
			9	19	28		
			(5)	(27)	(43)		
X̄			7.16	23.17	36.92	16.17	18.89%
F-test			**	**	**		
r _{xy}			-	0.036*	0.18*		
			-	-	0.13*		

* p > 0.05

** p < 0.05

จากตารางที่ 7 ปรากฏว่าโดยเฉลี่ยแล้วการวิเคราะห์หาความลำเอียงต่อภาคภูมิศาสตร์ของผู้สอบภาษาอังกฤษชุด กข. และ กขค. โดย วิธีการตอบสนองของข้อทดสอบชนิด 3 พารามิเตอร์ (Three-parameter Logistic Method: TPLM) เมื่อวิเคราะห์หาข้อทดสอบที่มีความลำเอียงสูงกว่าระดับน้อย ($0 > 0.40$) พบข้อทดสอบที่มีความลำเอียงจำนวนมากที่สุด คือ ประมาณ 36 ข้อต่อฉบับ รองลงมาได้แก่วิธีทดสอบความแตกต่างของสัดส่วนระหว่างกลุ่มด้วยไคสแควร์ (Chi-square Method: CSM) ซึ่งพบว่าแบบทดสอบทั้ง 6 ฉบับ ใน 3 ปี โดยเฉลี่ยแล้วมีความลำเอียงฉบับละประมาณ 38 ข้อ ทั้ง 2 วิธีนี้พบข้อทดสอบที่มีความลำเอียงต่อภาคจำนวนแตกต่างกันอย่างมีนัยสำคัญ ส่วนวิธีกำหนดจุดเดลต้า (Delta-Plot Method: DPM) พบข้อทดสอบที่มีความลำเอียงต่อภาคภูมิศาสตร์ของผู้สอบจำนวนน้อยที่สุด คือประมาณฉบับละ 7-8 ข้อ หรือประมาณ 4-5 เท่าของทั้ง 2 วิธีที่กล่าวแล้ว และแตกต่างกันอย่างมีนัยสำคัญ นอกจากนี้ยังมีข้อทดสอบประมาณร้อยละ 18-19 ของข้อทดสอบทั้งหมดที่การวิเคราะห์แต่ละวิธีพบว่ามีความลำเอียงต่อภาค และเมื่อทำการทดสอบด้วยวิธีอื่น ซึ่งอาจได้แก่อีก 1 หรือ 2 วิธี พบว่าข้อทดสอบเหล่านั้นมีความลำเอียงต่อภาคด้วย

อนึ่ง ยังพบว่าความสัมพันธ์ระหว่างจำนวนข้อทดสอบที่ลำเอียงต่อภาคของแบบทดสอบแต่ละฉบับในแต่ละปีของทั้ง 3 วิธีต่ำมาก ($r_{xy} = 0.036-0.18$) และไม่มีนัยสำคัญ ($p = 0.05$)

ดังนั้น จากตารางที่ 4-5, 5-9 และ 12-13 ทำให้สามารถกล่าวได้ว่า "แบบทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย ชุด กข. และ กขค. ปี พ.ศ. 2531-2533 มีความลำเอียงต่อเพศของผู้สอบ" ซึ่งสอดคล้องกับสมมติฐานการวิจัยข้อที่ 1 และจากตารางที่ 6-7, 10-11 และ 14-15 ทำให้สามารถกล่าวได้ว่า "แบบทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย ชุด กข. และ กขค. ปี พ.ศ. 2531-2533 มีความลำเอียงต่อภาคภูมิศาสตร์ของผู้สอบ" ซึ่งสอดคล้องกับสมมติฐานการวิจัยข้อที่ 2

นอกจากนี้ จากตารางที่ 16 และ 17 ทำให้สามารถกล่าวได้ว่า "การวิเคราะห์หาความลำเอียงของข้อทดสอบ 3 วิธีคือ วิธีกำหนดจุดเดลต้า วิธีทดสอบความแตกต่างของสัดส่วนของแต่ละกลุ่มด้วยไคสแควร์ และวิธีการตอบสนองของข้อทดสอบชนิด 3 พารามิเตอร์ ทำให้ได้จำนวนข้อทดสอบที่มีความลำเอียงแตกต่างกัน" ซึ่งสอดคล้องกับสมมติฐานการวิจัยข้อที่ 3

บทที่ 5

สรุปผล อภิปรายผล และข้อเสนอแนะ

ความเป็นมาและความสำคัญของปัญหา

ในแต่ละปีการศึกษา มีผู้สมัครสอบเข้าศึกษาต่อในมหาวิทยาลัยของรัฐและเอกชนจำนวนประมาณ 30,000 คน แต่จะมีจำนวนเพียงประมาณร้อยละ 25 เท่านั้นที่สามารถสอบเข้าศึกษาในมหาวิทยาลัยรับจำกัดของรัฐได้ การที่ผู้สมัครสอบไม่สามารถสอบเข้าศึกษาต่อได้นั้น ส่วนหนึ่งเป็นเพราะมีความรู้หรือความสามารถไม่เพียงพอ เมื่อเทียบกับผู้อื่นที่สามารถสอบเข้าศึกษาต่อได้ แต่อีกส่วนหนึ่งที่สำคัญมากเกิดจากความบกพร่องของคุณภาพของข้อทดสอบที่นำมาใช้ในการทดสอบ โดยเฉพาะอย่างยิ่งวิชาภาษาอังกฤษ เพราะจากการวิเคราะห์ข้อทดสอบวิชาภาษาอังกฤษ ชุด กข. และ กขค. ของผู้รู้ทางการวัดและประเมินผล รวมทั้งทางด้านหลักสูตรในหลายปีที่ผ่านมา พบว่าแบบทดสอบดังกล่าวยังมีข้อบกพร่องมาก เช่น ในปี พ.ศ. 2533 Alec Bradford (Bradford, 1933: 17) พบว่าแบบทดสอบภาษาอังกฤษชุด กข. มีปัญหาเกี่ยวกับคำตอบ 20 ข้อหรือร้อยละ 20 เช่น ไม่มีคำตอบที่ถูกต้องและมีคำตอบมากกว่า 1 ตัวเลือก เป็นต้น และแบบทดสอบภาษาอังกฤษชุด กขค. ก็มีปัญหาเกี่ยวกับคำตอบ 18 ข้อ หรือร้อยละ 18 ในปี พ.ศ. 2527 สุนันท์ สุกมลสันต์ (2527: 101-104) พบว่าแบบทดสอบภาษาอังกฤษชุด กข. และ กขค. มีปัญหาเกี่ยวกับคำตอบที่ถูกต้องร้อยละ 4 และ 18 และในปี 2528 สุนันท์ สุกมลสันต์ (2528: 1-20) ก็ยังพบว่าแบบทดสอบภาษาอังกฤษ ชุด กข. และ กขค. มีปัญหาเกี่ยวกับคำตอบที่ถูกต้องร้อยละ 15 และ 20 ตามลำดับอีกด้วย

ดังนั้น จะเห็นได้ว่าในแต่ละปีแบบทดสอบภาษาอังกฤษเข้ามหาวิทยาลัยมักมีปัญหาเกี่ยวกับคำตอบที่ถูกต้องประมาณชุดละ 20 ข้อ หรือร้อยละ 20 ทั้งนี้ยังไม่ได้รวมกับข้อบกพร่องด้านอื่นอีกที่มีผู้ศึกษาไว้เช่นเดียวกัน เช่นความไม่สอดคล้องของหลักสูตรระดับมัธยมศึกษากับเนื้อหาของแบบทดสอบเข้ามหาวิทยาลัย และความยากของข้อทดสอบ เป็นต้น ซึ่งข้อบกพร่องต่าง ๆ เหล่านี้ย่อมมีผลโดยตรงต่อผลการสอบเข้ามหาวิทยาลัยของผู้ที่สอบได้และสอบตก กล่าวโดยสรุปก็คือทำให้ผลการตัดสินได้-ตก คลาดเคลื่อนได้ ความบกพร่องต่าง ๆ ดังกล่าวแล้ว ทำให้ผู้วิจัยสงสัยว่าข้อทดสอบอาจมีความบกพร่องด้านอื่นอีกด้วย เช่นความลำเอียงของข้อทดสอบ (item bias) จึงมีความสนใจทำการศึกษารื่องนี้เพื่อเสนอผลใหม่ที่เกี่ยวข้องทราบ สำหรับใช้เป็นแนวทางแก้ปัญหาต่อไป

วัตถุประสงค์ของการวิจัย

1. เพื่อวิเคราะห์หาความลำเอียงต่อเพศและต่อภาคภูมิศาสตร์ของข้อทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย ปี พ.ศ. 2531-2533 โดย
 1. วิธีกำหนดจุดเดลต้า (Delta Plot Method)
 2. วิธีทดสอบความแตกต่างของสัดส่วนของแต่ละกลุ่มด้วยไคสแควร์ (Chi-square Method)
 3. วิธีการตอบสนองของข้อทดสอบชนิด 3 พารามิเตอร์ (Three-parameter Logistic Method)
2. เพื่อเปรียบเทียบจำนวนข้อทดสอบที่มีความลำเอียงต่อเพศ และต่อภาคภูมิศาสตร์ เมื่อวิเคราะห์ด้วย 3 วิธีดังกล่าว

สมมติฐานของการวิจัย

1. ข้อทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย ชุด กข. และ กขค. ปี พ.ศ. 2531-2533 น่าจะมีความลำเอียงต่อเพศของผู้สอบ
2. ข้อทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย ชุด กข. และ กขค. ปี พ.ศ. 2531-2533 น่าจะมีความลำเอียงต่อภาคภูมิศาสตร์ของผู้สอบ
3. วิธีการทดสอบหาความลำเอียงของข้อทดสอบ 3 วิธี คือวิธีกำหนดจุดเดลต้า วิธีทดสอบความแตกต่างของสัดส่วนของแต่ละกลุ่มด้วยไคสแควร์ และวิธีการตอบสนองของข้อทดสอบชนิด 3 พารามิเตอร์ น่าจะทำให้ได้จำนวนข้อทดสอบที่มีความลำเอียงต่อเพศ และต่อภาคภูมิศาสตร์ของผู้สอบแตกต่างกัน

ขอบเขตการวิจัย

1. วิเคราะห์หาความลำเอียงของข้อทดสอบของแบบทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย ชุด กข. และ กขค. ปี พ.ศ. 2531-2533 รวม 6 ฉบับ และมีข้อสอบทั้งหมด 600 ข้อ
2. วิเคราะห์หาความลำเอียงของข้อสอบเพียง 3 วิธีดังกล่าวแล้วเท่านั้น
3. ใช้ผลการสอบของผู้สอบจริง ทั้งผู้ที่สอบเข้าศึกษาต่อในมหาวิทยาลัย ได้และเข้าศึกษาต่อไม่ได้

ความจำกัดของการวิจัย

1. คำตอบของข้อทดสอบบางข้อมีปัญหาเช่น ไม่มีคำตอบที่ถูกต้อง หรือมีคำตอบที่ถูกต้องมากกว่า 1 คำตอบ
2. ผู้วิจัย ไม่มีคีย์เฉลย (keys) อย่างเป็นทางการของข้อทดสอบแต่ละข้อ
3. แนวคิดและความหมายของคำว่า "ความลำเอียงของข้อทดสอบ (item bias)" ของการวิเคราะห์หาความลำเอียงแต่ละวิธีแตกต่างกัน
4. ศึกษาความลำเอียงต่อภาคภูมิศาสตร์ โดยการเปรียบเทียบผลการสอบระหว่างผู้สอบจากภาคกลางกับภาคอื่น ๆ อีก 4 ภาคเท่านั้น ไม่ได้เปรียบเทียบระหว่างภาคอื่นอีก
5. ไม่ได้มุ่งศึกษาถึงสาเหตุหรือแหล่งที่เกิดของความลำเอียงของข้อทดสอบแต่ละข้อ เพราะข้อทดสอบมีจำนวนมากและต้องศึกษาในหลายลักษณะ (เช่น คำนวณแล้วมีอย่างน้อยประมาณ $5 \times 100 \times 6 = 3,000$ กราน!) ๑

ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อเป็นข้อมูลให้แก่คณะกรรมการหรือผู้ที่รับผิดชอบในการสร้างแบบทดสอบคัดเลือกเข้ามหาวิทยาลัยปีต่อ ๆ ไป ได้ตระหนักถึงข้อบกพร่องของข้อทดสอบด้านความลำเอียงต่อเพศ และภาคภูมิศาสตร์ของผู้สอบ เพื่อหาทางปรับปรุงแก้ไข หรือป้องกันไม่ให้เกิดข้อบกพร่องดังกล่าว เพื่อให้เกิดความเป็นธรรมแก่ผู้สอบและเป็นผลดีต่อระบบการศึกษาในส่วนรวมด้วย
2. เพื่อเป็นการตรวจสอบคุณภาพของแบบทดสอบภาษาอังกฤษเข้ามหาวิทยาลัยในระยะ 3 ปีที่แล้ว จำนวน 6 ฉบับ อีกวิธีหนึ่งว่าแบบทดสอบดังกล่าวมีคุณภาพเป็นอย่างไร
3. เพื่อเป็นการแนะนำการวิเคราะห์หาคุณภาพของข้อทดสอบอีกวิธีหนึ่ง ให้แก่นักทดสอบและผู้ที่เกี่ยวข้องหรือสนใจทั่วไป ให้เห็นถึงความสำคัญ แนวคิด และวิธีการจัดการกระทำ โดยมุ่งหวังว่าจะเป็นวิธีแพร่หลายต่อไป เพื่อให้การสร้างและพัฒนาข้อสอบในประเทศไทยมีความก้าวหน้ายิ่งขึ้น

วิธีดำเนินการ

1. ประชากร

ประชากรการวิจัยครั้งนี้ ได้แก่ผู้สอบแบบทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย ชุด กข. และ/หรือ กขค. ปี พ.ศ. 2531-2533 จำนวน 6 กลุ่ม ๆ ละ ประมาณ 30,000 - 80,000 คน แต่ละกลุ่มแบ่งตามเพศของผู้สอบได้ 2 กลุ่ม คือ เพศชาย และเพศหญิง ซึ่งมีประมาณกลุ่มละ 8,000 - 40,000 คน และแต่ละกลุ่มแบ่งตามภาคภูมิศาสตร์ของผู้สอบได้อีก 5 กลุ่ม ซึ่งมีประมาณกลุ่มละ 1,000 - 47,000 คน ตามจำนวนผู้ที่เข้าสอบจริง

2. ผลวิจัย

ผลวิจัยได้แก่ตัวอย่างของประชากรของแต่ละกลุ่ม ซึ่งได้มาจากการสุ่มอย่างง่ายโดยอาศัยเครื่องคอมพิวเตอร์ช่วยในการสุ่มตามจำนวนสูงสุดเท่าที่โปรแกรมคอมพิวเตอร์สำหรับวิเคราะห์ข้อทดสอบจะสามารถทำงานได้ (ไม่เกินกลุ่มละ 3,000 คน) และแต่ละกลุ่มมีจำนวนผู้สอบเพศชายและเพศหญิงเท่ากัน ทำให้ผลวิจัยแต่ละกลุ่มมีขนาดระหว่าง 424-3,000 คน ซึ่งมีขนาดมากกว่าขนาดกลุ่มตัวอย่างที่เหมาะสมเพื่อการวิจัย เมื่อกำหนดให้มีความเชื่อมั่นในการสุ่มอย่างง่ายเท่ากับร้อยละ 99 และความคลาดเคลื่อนในการสุ่มไม่เกินร้อยละ 5 จึงเชื่อได้ว่าเป็นผลวิจัยที่มีขนาดเพียงพอ และเป็นตัวอย่างที่ดีของประชากร

3. เครื่องมือที่ใช้ในการวิจัย

ได้แก่ ผลการสอบรายชื่อของข้อทดสอบจำนวน 600 ข้อ จากแบบทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย ชุด กข. และ กขค. ปี พ.ศ. 2531-2533 ข้อทดสอบดังกล่าวสร้างขึ้น โดยคณะกรรมการสร้างแบบทดสอบของทบวงมหาวิทยาลัย ข้อทดสอบเหล่านี้เป็นแบบปรนัย ชนิดเลือกตอบ และมี 4 ตัวเลือก

4. การเก็บรวบรวมข้อมูล

- เขียนจดหมายติดต่ออย่างเป็นทางการ เพื่อขอสัมภาระแม่เหล็กบันทึกข้อมูลคอมพิวเตอร์ ที่บันทึกผลการสอบรายชื่อของวิชา 05 (ภาษาอังกฤษชุด กข.) และ 10 (ภาษาอังกฤษ ชุด กขค.) ประจำปี 2531-2533 จากทบวงมหาวิทยาลัย
- ดำเนินการสุ่มเลือกผลวิจัยตามขนาดและเงื่อนไขที่ต้องการโดยใช้โปรแกรมคอมพิวเตอร์

5. การวิเคราะห์ข้อมูล

1. ใช้โปรแกรมคอมพิวเตอร์ชื่อ SPSSX (The Statistical Packages for the Social Sciences, Version X) ของสถาบันบริการคอมพิวเตอร์ของจุฬาลงกรณ์มหาวิทยาลัย เช่น ใช้ sub-program ชื่อ SAMPLE เพื่อสุ่มตัวอย่างของประชากร ใช้ FACTOR เพื่อวิเคราะห์หาความมีมิติเดียวของข้อทดสอบ และใช้ CROSSTABS เพื่อวิเคราะห์หาความลำเอียง โดยวิธีใช้ Chi-square เป็นต้น

2. ใช้โปรแกรมคอมพิวเตอร์ชื่อ ITEMX เพื่อวิเคราะห์ข้อทดสอบรายข้อตามวิธีประเพณีนิยม และใช้โปรแกรมชื่อ LOGLIST เพื่อวิเคราะห์ข้อทดสอบรายข้อตามทฤษฎีการตอบสนองของข้อทดสอบ (Item Response Theory)

3. ใช้โปรแกรมที่ผู้วิจัยเขียนขึ้นคือ BIAS เพื่อวิเคราะห์หาความลำเอียงของข้อทดสอบตามวิธีที่ 1 และ ICCTEST เพื่อวิเคราะห์หาความลำเอียงของข้อทดสอบตามวิธีที่ 3

6. สถิติที่ใช้

ใช้สถิติที่เกี่ยวข้องหลายอย่าง เช่น สถิติบรรยาย การสุ่มอย่างง่าย Factor Analysis, Chi-square tests, Classical Item Analysis, Logistic Item Analysis, ICC Test, Bias Test, F-test และ Simple Correlation เป็นต้น

ผลการวิจัย

1. ข้อทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย ชุด กข. และ กขค. ปี พ.ศ. 2531-2533 มีความลำเอียงต่อเพศ โดยชุด กข. มีข้อทดสอบที่มีความลำเอียงต่อเพศโดยเฉลี่ยประมาณ 7-28 ข้อ (ร้อยละ 7-28) แล้วแต่ผลการวิเคราะห์แต่ละวิธี และชุด กขค. มีข้อทดสอบที่ลำเอียงต่อเพศโดยเฉลี่ยประมาณ 4-41 ข้อ (ร้อยละ 4-41)
2. ข้อทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย ชุด กข. และ กขค. ปี พ.ศ. 2531-2533 มีความลำเอียงต่อภาคภูมิศาสตร์ของผู้สอบ โดยชุด กข. มีข้อทดสอบที่มีความลำเอียงต่อภาคโดยเฉลี่ยประมาณ 6-45 ข้อ (ร้อยละ 6-45) แล้วแต่ผลการวิเคราะห์แต่ละวิธี และชุด กขค. มีข้อทดสอบที่ลำเอียงต่อภาคโดยเฉลี่ยประมาณ 5-43 ข้อ (ร้อยละ 5-43)

3. ข้อทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย ชุด กข. และ กขค. ปี พ.ศ. 2531-2533 มีความลำเอียงต่อผู้สอบจากภาคอื่นมากกว่าจากภาคกลางประมาณ 2-3 เท่า และมีแนวโน้มว่ามีความลำเอียงต่อผู้สอบเพศชายมากกว่าเพศหญิง
4. การวิเคราะห์หาความลำเอียงของข้อทดสอบด้วย 3 วิธีที่แตกต่างกัน พบข้อทดสอบที่มีความลำเอียงจำนวนแตกต่างกันอย่างมีนัยสำคัญ และสัมพันธ์กันอย่างไม่มีนัยสำคัญ การวิเคราะห์โดย วิธีการตอบสนองของข้อทดสอบชนิด 3 พารามิเตอร์ (Three-parameter Logistic Method) เมื่อวิเคราะห์หาเฉพาะข้อทดสอบที่ลำเอียงในระดับปานกลางและมากเท่านี้พบข้อทดสอบที่ลำเอียงจำนวนมากที่สุด รองลงมาได้แก่ วิธีทดสอบความแตกต่างของสัดส่วนของแต่ละกลุ่มด้วย ไคสแควร์ (Chi-square Method) และวิธีกำหนดจุดเดลต้า (Delta Plot Method) พบข้อทดสอบที่ลำเอียงจำนวนน้อยที่สุด และน้อยกว่าจำนวนของวิธีที่ 1-2 ดังกล่าวแล้วประมาณ 3-4 เท่า

อภิปรายผล

ผลจากการวิจัยครั้งนี้มีประเด็นที่น่าสนใจหลายอย่างดังนี้

1. เหตุใดข้อทดสอบภาษาอังกฤษเข้ามหาวิทยาลัยจึงมีความลำเอียงต่อเพศชายมากกว่าเพศหญิง
ผลของการวิจัยจากตารางที่ 4-5, 8-9 และ 12-13 เกี่ยวกับจำนวนข้อทดสอบที่ลำเอียงต่อเพศเมื่อวิเคราะห์ด้วยทั้ง 3 วิธีสรุปได้ดังนี้

จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 18 จำนวนข้อทดสอบที่มีความลำเอียงต่อเพศ เมื่อวิเคราะห์ด้วย 3 วิธี

ปี พ.ศ.	แบบทดสอบชุด	DPM		CSM		TPLM*	
		M	F	M	F	M	F
2531	กข	1	2	9	6	n/a	n/a
2532	กข	7	6	13	14	n/a	n/a
2533	กข	1	3	5	4	n/a	n/a
2531	กขค	3	3	15	7	n/a	n/a
2532	กขค	2	2	12	5	n/a	n/a
2533	กขค	1	1	9	5	n/a	n/a
รวม	M = 78, F = 58 $\bar{X} = 6.50, 4.83$ t = 0.8885**	15	17	63	41	n/a	n/a

* ค่า ϕ ไม่อาจบอกได้แน่ชัดว่าข้อสอบลำเอียงต่อกลุ่มใดและ n/a = not applicable

** $p > 0.05$

จากตารางที่ 18 ปรากฏว่าผลรวมของข้อทดสอบที่ลำเอียงต่อเพศชายมีมากกว่าต่อเพศหญิง คือ 78:58 แต่ค่าเฉลี่ยต่อฉบับต่อวิธีของข้อทดสอบที่ลำเอียงต่อเพศชายและหญิง คือ 6.50 และ 4.83 แตกต่างกันอย่างไม่มีนัยสำคัญ ($p = 0.05$) แสดงว่าจากข้อมูลที่ปรากฏในตารางที่ 18 ยิ่งไม่อาจกล่าวได้ว่าแบบทดสอบภาษาอังกฤษชุด กข. และ กขค. ปี 2531-2533 โดยเฉลี่ยแล้ว มีข้อทดสอบที่ลำเอียงต่อเพศใดเพศหนึ่งโดยเฉพาะ เพียงแต่มีแนวโน้มว่าข้อทดสอบแต่ละชุดมีความลำเอียงต่อผู้สอบเพศชายมากกว่าเพศหญิง ซึ่งอาจเป็นเพราะว่า

1. โดยปกติทั่วไปแล้ว นักเรียน นิสิต หรือนักศึกษาเพศหญิง เป็นผู้มีความรู้ความสามารถทางภาษาสูงกว่าผู้เรียนเพศชาย ทั้งนี้จากผลการวิจัยทั้งภายในและต่างประเทศจำนวนมากต่างก็ค้นพบปรากฏการณ์ดังกล่าวแล้ว (เช่น Burstall, 1978: 3-8); ไกรคุง อนันตกุล, 2525: 54; ประทุม เจริญผล, 2526: 81)
2. ข้อทดสอบจำนวนหนึ่งอาจมีความลำเอียงอันเกิดจากการเลือกเนื้อหาอย่างลำเอียง (bias in selection) ของผู้สร้างข้อทดสอบด้วยการเลือกเนื้อหาที่ไม่เป็นสัดส่วนที่ดีจากหลักสูตรและเนื้อหาที่เลือกมาเอื้อประโยชน์ให้แก่ผู้สอบที่เป็นเพศหญิงมากกว่าเพศชาย รวมทั้งข้อทดสอบอาจมีความลำเอียงที่เกิดจากการสร้างข้อทดสอบแต่ละข้อ (bias in construction) โดยผู้สร้างอาจใช้คำศัพท์ โครงสร้าง และสำนวนภาษาจำนวนหนึ่งซึ่งเป็นที่ยอมรับกันดีเฉพาะผู้สอบเพศหญิง หรือเป็นสิ่งที่ผู้สอบเพศหญิงมีความสนใจมากกว่าผู้สอบเพศชายก็ได้ (Shepard, 1982: 10) ผู้ที่จะตอบข้อสงสัยเหล่านี้ได้จะต้องทำการศึกษา วิเคราะห์ และพิจารณาเนื้อหาในเชิงตรรก (Logical Analysis) เป็นรายข้ออย่างจริงจัง ประกอบด้วยความรู้ที่ได้จากการวิเคราะห์เชิงสถิติ (Statistic Analysis) ที่ได้จากงานวิจัยเช่นนี้ ทั้งนี้เพราะความลำเอียงของข้อทดสอบเป็นสิ่งที่เกิดขึ้นในข้อทดสอบ

แต่อย่างไรก็ตามเหตุผลทั้ง 2 ประการดังกล่าวแล้วค่อนข้างขัดแย้งกัน และยังไม่อาจเชื่อได้แน่ว่าเหตุผลใด มีน้ำหนักความน่าเชื่อถือได้มากกว่ากัน เนื่องจากแบบทดสอบทั่ว ๆ ไปที่ใช้ในการเปรียบเทียบผลสัมฤทธิ์ในการเรียนภาษา (Language learning achievement) หรือสมรรถภาพทางภาษา (Language proficiency) อาจยังไม่ได้คำนึงถึงเรื่องความลำเอียงต่อเพศของข้อทดสอบ เนื่องจากเรื่องความลำเอียงของข้อทดสอบเป็นแนวคิดที่นักทดสอบเพิ่งนำมาปฏิบัติเมื่อไม่นานมานี้ คือเมื่อประมาณ ค.ศ. 1970 นี้เอง (Berk, 1983: 1) ดังนั้นแบบทดสอบที่ใช้ในงานวิจัย เพื่อการเปรียบเทียบดังกล่าวอาจประกอบด้วยข้อทดสอบที่มีความลำเอียงต่อเพศก็ได้ ทำให้ผลการวิจัยยังไม่อาจเชื่อได้มากนัก ดังนั้นข้อค้นพบในอดีตในประเด็นที่ 1 จึงยังเป็นเรื่องที่น่าค้นหาคำตอบต่อไป

2. เหตุใดข้อทดสอบภาษาอังกฤษเข้ามหาวิทยาลัยจึงมีความลำเอียงต่อผู้สอบจากภาคภูมิศาสตร์อื่นมากกว่าต่อภาคกลาง

ผลของการวิจัยจากตารางที่ 6-7, 10-11 และ 14-15 เกี่ยวกับจำนวนข้อทดสอบที่ลำเอียงต่อผู้สอบจากภาคภูมิศาสตร์ต่าง ๆ เมื่อวิเคราะห์ด้วย 3 วิธีการสรุปได้ดังนี้

ตารางที่ 19 จำนวนข้อทดสอบที่ล่าช้าเชิงต่อภาคภูมิศาสตร์เมื่อวิเคราะห์ด้วย 3 วิธี

วิธีวิเคราะห์	ปี พ.ศ.	แบบทดสอบชุด	= C	E	NE	N	S
DPM	2531	กข	3	3	9	3	4
	2532	กข	5	6	6	7	7
	2533	กข	2	4	3	4	5
	2531	กขค	3	2	3	4	1
	2532	กขค	1	2	3	3	2
	2533	กขค	2	6	16	7	8
CSM	2531	กข	11	32	21	6	24
	2532	กข	9	43	35	8	34
	2533	กข	11	31	20	15	27
	2531	กขค	2	13	18	22	26
	2532	กขค	6	15	25	25	23
	2533	กขค	3	15	14	19	16
TPLM**	2531-2533	กข และ กขค	n/a	n/a	n/a	n/a	n/a
\bar{X}			4.85	14.33	14.42	10.25	14.75
F-test			-	*	*	*	*

* $p > 0.05$

** ไม่อาจบอกได้แน่ชัด

จากตารางที่ 19 ปรากฏว่าโดยเฉลี่ยแล้ว ผลการวิเคราะห์หาข้อทดสอบที่มีความ
ลำเอียงต่อภาคภูมิศาสตร์ 2 วิธีคือ Delta Plot Method และ Chi-square Method พบว่า
จำนวนข้อทดสอบที่มีความลำเอียงต่อผู้สอบจากภาคตะวันออก (E) ภาคตะวันออกเฉียงเหนือ (NE)
ภาคเหนือ (N) และภาคใต้ (S) มีจำนวนใกล้เคียงกันคือประมาณ 10-15 ข้อ และแตกต่างกัน
อย่างไม่มีนัยสำคัญ แต่ต่างก็มีค่ามากกว่าข้อทดสอบที่ลำเอียงต่อผู้สอบจากภาคกลางซึ่งมีประมาณ
5 ข้ออย่างมีนัยสำคัญ ซึ่งหมายความว่าโดยเฉลี่ยแล้วแบบทดสอบภาษาอังกฤษเข้ามหาวิทยาลัยมีความ
ลำเอียงต่อผู้สอบจากภาคต่าง ๆ มากกว่าต่อผู้สอบจากภาคกลางประมาณ 2-3 เท่า ปรากฏการณ์
ดังกล่าวนี้อาจเป็นเพราะว่า

1. ความไม่เท่าเทียมกันด้านโอกาสทางการศึกษาของผู้สอบจากภาคกลาง และภาค
อื่น ๆ ซึ่งโดยมากยังมีสภาพเป็นชนบท ขาดความพร้อมทางอุปกรณ์การเรียน
ขาดครู-อาจารย์ผู้สอนที่มีคุณวุฒิทางการศึกษาสูง ขาดโอกาสในการใช้หรือรับรู้สื่อ
ภาษาอังกฤษ รวมทั้งความขาดแคลนอื่น ๆ เมื่อเทียบกับผู้สอบที่อยู่ในภาคกลาง
ซึ่งโดยมากอยู่ในสังคมเมือง จึงมีโอกาสดูและมีความพร้อมในด้านต่าง ๆ ดังกล่าวแล้ว
ดีกว่า ความแตกต่างดังกล่าวนี้คงเป็นปัจจัยที่ลำเอียงอย่างหนึ่งที่ทำให้สัมฤทธิ์ผลในการ
เรียนภาษาอังกฤษของผู้เรียนในต่างจังหวัดต่ำกว่าผู้เรียนในกรุงเทพมหานคร
(สำนักงานคณะกรรมการการศึกษาแห่งชาติ, 2531) เช่น ในปี พ.ศ.
2516 อ้าไพ อนันต์ศฤงคาร (2516: 41) ศึกษาพบว่านักเรียนชั้นมัธยมศึกษา
ปีที่ 3 ในกรุงเทพมหานครมีผลสัมฤทธิ์ในการเรียนภาษาอังกฤษสูงกว่านักเรียน
จากจังหวัดสงขลาอย่างมีนัยสำคัญ และในปี 2533 เฉลิมชัย มิติภักษ์ (2532:
39-42) ศึกษาว่าความสามารถในการฟังภาษาอังกฤษของนักเรียนชั้นมัธยม-
ศึกษาปีที่ 3 ที่เรียนอยู่ในโรงเรียนในเขตอำเภอเมืองนครราชสีมา สูงกว่าของ
นักเรียนที่อยู่นอกเขตอำเภอเมืองอย่างมีนัยสำคัญ เป็นต้น
2. ข้อทดสอบจำนวนหนึ่งอาจมีความลำเอียงอันเกิดจากการเลือกเนื้อหาอย่างลำเอียง
(bias in selection) ของผู้สร้างข้อทดสอบด้วยการเลือกเนื้อหาที่ไม่
เป็นสัดส่วนที่ดีจากหลักสูตร และข้อทดสอบอีกจำนวนหนึ่งอาจมีความลำเอียงจาก
การสร้างข้อทดสอบของกรรมการสร้างข้อทดสอบ (bias in construction)
โดยผู้สร้างอาจใช้คำศัพท์ เนื้อเรื่องสำหรับอ่าน (reading passage) โครง-
สร้าง และสำนวนภาษาจำนวนหนึ่งที่เป็นที่รู้จักกันดีในระหว่างผู้สอบในภาคกลาง
แต่ไม่เป็นที่รู้จักกันทั่วไป สำหรับผู้สอบจากภาคอื่น ๆ ความผิดพลาดนี้อาจเกิดจาก
ความไม่ตั้งใจของผู้สร้างข้อทดสอบ แต่สิ่งเหล่านี้ล้วนแต่มีอิทธิพลต่อโครงสร้าง
ความรู้ (schema) ของผู้สอบ และโครงสร้างความรู้ก็มีอิทธิพลต่อความเข้าใจ

เรื่องนี้อ่านอย่างมีนัยสำคัญด้วย (Mandler, 1979: 263-6; Carrell, 1984: 332-342) ดังนั้นหากผู้สอบอยู่ในสังคมเมืองก็ยังมีโอกาสได้เรียนรู้สิ่งต่าง ๆ รอบตัวมากกว่า ทั้งจากความพร้อมด้านโอกาสทางการศึกษาเชิงรูปนัย (formal education) ที่ดีกว่าแล้ว ก็ยังมีโอกาสทางการศึกษาเชิงอรูปนัย (informal education) เช่นจากสื่อมวลชนต่าง ๆ ดีกว่าผู้สอบจากภาคอื่น ๆ ที่มีโอกาสน้อยกว่าด้วย

ดังนั้น จากเหตุผลที่สำคัญดังกล่าวแล้ว 2 ข้อข้างต้น อาจเป็นปัจจัยที่สำคัญที่ทำให้แบบทดสอบที่มีข้อทดสอบที่มีความลำเอียงต่อผู้สอบจากภาคอื่น ๆ มากกว่าจากภาคกลาง

3. เหตุใดจำนวนข้อทดสอบที่ลำเอียงซึ่งได้จากการวิเคราะห์หาความลำเอียงแต่ละวิธีจึงแตกต่างกัน

จากตารางที่ 16 และ 17 ซึ่งเป็นตารางสรุปเปรียบเทียบเกี่ยวกับจำนวนข้อทดสอบที่ลำเอียงซึ่งได้จากการวิเคราะห์ 3 วิธี ปรากฏว่าจำนวนข้อทดสอบที่ลำเอียงของแบบทดสอบแต่ละฉบับในแต่ละวิธีมีจำนวนแตกต่างกันและโดยเฉลี่ยแล้ว การวิเคราะห์โดยวิธีการตอบสนองของข้อทดสอบชนิด 3 พารามิเตอร์ (Three-parameter Logistic Method) พบข้อทดสอบที่มีความลำเอียงต่อเพศหรือต่อภาคภูมิศาสตร์จำนวนมากที่สุด รองลงมาได้แก่ผลของการวิเคราะห์ด้วยวิธีทดสอบความแตกต่างของสัดส่วนระหว่างกลุ่มด้วย ไคสแควร์ (Chi-square Method) ซึ่งได้ข้อทดสอบที่ลำเอียงจำนวนแตกต่างกันอย่างมีนัยสำคัญ ส่วนวิธีกำหนดจุดเดลต้า (Delta Plot Method) พบข้อทดสอบที่ลำเอียงน้อยที่สุด ประมาณ 4-5 เท่าของ 2 วิธีที่กล่าวแล้วและจำนวนต่างกันอย่างมีนัยสำคัญ

ความแตกต่างของจำนวนข้อทดสอบที่ลำเอียงดังกล่าวอาจเกิดจากสาเหตุต่อไปนี้คือ

1. แนวคิดเกี่ยวกับความลำเอียงของข้อทดสอบของแต่ละวิธีแตกต่างกัน ทำให้แต่ละวิธีนิยามเชิงปฏิบัติการ (operation definition) ของความลำเอียงของข้อทดสอบแตกต่าง จึงทำการคำนวณหาความลำเอียงด้วยวิธีการและสูตรที่ต่างกัน จนเป็นเหตุให้จำนวนข้อทดสอบที่ลำเอียงต่างกัน เช่น (Shepard, 1982: 23; Osterlind, 1983: 55):

1. วิธีกำหนดจุดเดลต้า เชื่อว่า

ข้อทดสอบที่ลำเอียงคือ ข้อทดสอบที่มีค่าเดลต้าของผู้สอบแต่ละกลุ่ม เรียงลำดับแตกต่างกัน

2. วิธีทดสอบความแตกต่างของสัดส่วนระหว่างกลุ่มด้วย ไคสแควร์
เชื่อว่า

ข้อทดสอบที่ลำเอียงคือข้อทดสอบซึ่งผู้สอบที่ได้คะแนนใน
ช่วงระดับคะแนนเดียวกัน มีสัดส่วนของการตอบข้อ
ทดสอบถูกและผิดแตกต่างกัน

3. วิธีการตอบสนองของข้อทดสอบชนิด 3 พารามิเตอร์ เชื่อว่า

ข้อทดสอบที่ลำเอียง คือข้อทดสอบซึ่งผู้สอบที่มีความ
สามารถเท่ากันมีโอกาสตอบข้อทดสอบได้ถูกต้องแตกต่างกัน

ดังนั้น จะเห็นได้ว่าวิธีที่ 1 มุ่งพิจารณาที่ระดับความยาก-ง่ายของ
ข้อทดสอบแต่ละข้อ เป็นสำคัญ เนื่องจากคำนวณหาความลำเอียงของข้อทดสอบ
โดยอาศัยค่าเดลต้า (Δ) ซึ่งเป็นค่าที่แสดงความยาก-ง่ายของข้อทดสอบที่
แปลงมาจากค่าอัตราส่วนของการตอบข้อทดสอบแต่ละข้อถูก (p) มากน้อยเพียง
ใด ซึ่งได้แก่ค่าดัชนีของความยากง่ายของข้อทดสอบโดยตรง

ส่วนวิธีที่ 2 มุ่งพิจารณาที่ระดับความสามารถของกลุ่มผู้สอบและ
ความยากง่ายของข้อทดสอบแต่ละข้อ เป็นสิ่งสำคัญ จึงเป็นวิธีที่ซับซ้อนกว่าวิธีที่ 1
เล็กน้อยในเชิงของแนวคิดและวิธีการคำนวณ แต่ยังเป็นวิธีที่มีจุดอ่อนมากเกี่ยวกับการ
การแบ่งจำนวนกลุ่มของผู้สอบว่าควรจะเป็นกี่กลุ่มสำหรับการวิเคราะห์หาความ
ลำเอียงของข้อทดสอบแต่ละข้อ เนื่องจากยังไม่มีกฎเกณฑ์ที่แน่นอน (Osterlind,
1983: 41-42) และการที่ผู้สอบที่มีคะแนนรวมอยู่ในระดับเดียวกัน อาจมี ความ
สามารถที่แท้จริง แตกต่างกันได้ ทั้งนี้ขึ้นอยู่กับช่วงห่างของคะแนน (score
internal) และที่มาของคะแนนรวม

สำหรับวิธีที่ 3 นี้มุ่งพิจารณาความสามารถของผู้สอบแต่ละคน และ
พารามิเตอร์ของข้อทดสอบแต่ละข้ออย่างละเอียด คืออำนาจจำแนกของข้อทดสอบ
(a) ความยากง่ายของข้อทดสอบ (b) และค่าการเดาของข้อทดสอบ (c)
เป็นสำคัญ จึงเป็นวิธีที่ซับซ้อนมากที่สุดในเรื่องของการคำนวณ และเป็นวิธีที่ไว
(sensitive) มากที่สุด ในเรื่องของการค้นหาข้อทดสอบที่ลำเอียง หากพิจารณา
ถึงความลำเอียงในระดับน้อยด้วย (Osterlind, 1983: 77)

2. เกณฑ์ในการตัดสินความลำเอียงของข้อทดสอบแต่ละวิธีไม่แน่นอน เนื่องจากยัง
ขาดเกณฑ์กลางที่เป็นที่ยอมรับกันทั่วไป ดังนั้นงานวิจัยแต่ละเรื่อง จึงอาจใช้เกณฑ์
ที่แตกต่างกัน เช่น เกณฑ์ในการตัดสินระยะห่างตั้งฉากของค่าอันดับคู่เดลต้าไปยัง
แกนสำคัญ (d) นักวิจัยอาจใช้ $d = \pm 2S_d$, $\pm 3 S_d$, $\pm 0.75 z$ -score unit

หรือ ± 1.5 z-score unit ก็ได้ (Angoff, 1982: 107; Osterlind, 1983: 35) ทั้งนี้แล้ววัตถุประสงค์ของการวิจัยแต่ละเรื่อง หรือในกรณีของการวิเคราะห์วิธีที่ 2 (Chi-square Method) ผู้วิจัยอาจกำหนดระดับความมีนัยสำคัญของตนเองได้ว่าควรจะเป็น $\alpha = 0.05$ หรือ 0.01 รวมทั้งการกำหนดจำนวนกลุ่มของผู้สอบซึ่งได้กล่าวมาแล้วว่ายังไม่มีเกณฑ์ที่แน่นอน จำนวนกลุ่มของผู้สอบที่แบ่งตามระดับความสามารถโดยใช้คะแนนรวมนี้ มีผลโดยตรงต่อเกณฑ์ในการตัดสินความล่าเอียงของข้อทดสอบด้วย เนื่องจากสูตรทั่วไปในการคำนวณหาค่าระดับความมีนัยสำคัญ เมื่อคำนึงถึงค่าอัตราความคลาดเคลื่อน (error rate) แล้ว คือ $\alpha = \frac{1}{j} \alpha_c$ และ j คือจำนวนกลุ่ม ดังนั้น ถ้าหากว่าจำนวนกลุ่มเปลี่ยนไป เกณฑ์ในการตัดสินความล่าเอียงของข้อทดสอบก็เปลี่ยนไปด้วย

ส่วนวิธีวิเคราะห์ที่ 3 คือ Three-parameter Logistic Method นั้น แม้ว่ามีนักวิจัยบางท่านเสนอว่าค่า $\phi < 0.40 =$ ล่าเอียงน้อย ϕ ระหว่าง $0.40-0.70 =$ ล่าเอียงปานกลาง และ $\phi > 0.70 =$ ล่าเอียงมาก (Rudner, 1977 อ้างจากสรัสตีกดี อมรรัตนตีกดี, 2531: 44) แต่ไม่ได้ระบุไว้แน่ชัดว่าคำว่า "ล่าเอียงน้อย" นั้น มี "นัยสำคัญ" หรือไม่ ดังนั้นในการวิจัยครั้งนี้ผู้วิจัยจึงไม่ได้ค้นหาข้อทดสอบที่มีความล่าเอียงน้อย

ดังนั้นจะเห็นได้ว่าเกณฑ์ในการตัดสินความล่าเอียงของข้อทดสอบของการวิเคราะห์แต่ละวิธียังไม่เกณฑ์กลางที่ยอมรับกันทั่วไป หากเปลี่ยนแปลงเกณฑ์ของวิธีใดวิธีหนึ่ง ก็ย่อมทำให้จำนวนข้อทดสอบที่ล่าเอียงเปลี่ยนแปลงตามไปด้วย

ด้วยเหตุผลที่สำคัญดังกล่าวแล้ว 2 ประการข้างต้น น่าจะเป็นสาเหตุที่ทำให้ข้อทดสอบที่ล่าเอียงแต่ละวิธีแตกต่างกัน และเมื่อพิจารณาถึงความซับซ้อนของแนวคิดเกี่ยวกับความล่าเอียงของข้อทดสอบและวิธีการคำนวณแล้ว อาจกล่าวได้ว่าวิธีที่ 1 เป็นวิธีที่อนุรักษ์นิยม (conservative) มากที่สุด รองลงมาได้แก่ วิธีที่ 2 และวิธีที่ 3 เป็นวิธีที่ไว (sensitive) มากที่สุด เมื่อคำนึงถึงการค้นหาข้อทดสอบที่ล่าเอียงในระดับน้อย หรือต่ำด้วย ดังนั้น โดยสภาพทั่ว ๆ ไปแล้วจำนวนข้อทดสอบที่ล่าเอียงที่พบจากวิธีที่ 3 จึงมีจำนวนมากที่สุด (ถ้ารวมข้อที่ล่าเอียงน้อยด้วย) รองลงมาได้แก่ วิธีที่ 2 และวิธีที่ 1 พบข้อทดสอบที่ล่าเอียงจำนวนน้อยที่สุด ผลการวิจัยครั้งนี้จึงสอดคล้องกับสภาพทั่วไปดังกล่าวแล้ว และสอดคล้องกับผลการวิจัยของทัศนีย์ พิรมนตรี (2530, 73) ที่พบว่าข้อทดสอบที่ล่าเอียงต่อภาคภูมิศาสตร์ของวิชาคณิตศาสตร์ที่วิเคราะห์โดยวิธีที่ 3 มีมากกว่าวิธีที่ 1 (ส่วนวิธี Chi-square Method นั้น ทัศนีย์ พิรมนตรี ใช้วิธีที่ต่างจากวิธีที่ผู้วิจัยใช้ ในการวิจัยครั้งนี้)

และสอดคล้องกับผลการวิจัยของสُرศักดิ์ อมรัตน์ศักดิ์ (2531: 170-182) ในการเปรียบเทียบผลของการวิเคราะห์หาความลำเอียงของ 4 วิธีจากการวิเคราะห์แบบทดสอบคัดเลือก 4 ฉบับ

1. เหตุใดจำนวนข้อทดสอบที่ลำเอียงจากการวิเคราะห์แต่ละวิธีจึงมีความสัมพันธ์กัน
อย่างไม่มีนัยสำคัญ

ปรากฏการณ์ดังกล่าวนี้อาจเป็นเพราะว่า

1. การวิเคราะห์หาความลำเอียงแต่ละวิธีมุ่งวิเคราะห์ลักษณะของข้อทดสอบข้อเดียวกันคนละด้าน เนื่องจากมีแนวคิดเกี่ยวกับความลำเอียงแตกต่างกัน และวิธีการทดสอบแตกต่างกัน ตามรายละเอียดที่ได้กล่าวมาแล้วข้างต้นในประเด็นที่ 3
2. วิธีการวิเคราะห์หาความลำเอียงทั้ง 3 วิธียังมีข้อบกพร่องบางอย่าง เช่น
 - ก. ข้อทดสอบที่ลำเอียงที่พบจากวิธีที่ 1 มักเป็นข้อที่มีอำนาจจำแนกสูงหรือต่ำมาก นอกจากเป็นข้อทดสอบที่มีค่าความยากง่าย (Δ) ของกลุ่มแตกต่างกันมาก ดังนั้นหากข้อทดสอบมีค่าอำนาจจำแนกของกลุ่มแตกต่างกันมาก ค่าความลำเอียงของข้อทดสอบอาจเป็นค่าที่ผิดพลาด (Angoff, 1982: 104)
 - ข. ข้อทดสอบที่ลำเอียงที่พบจากวิธีที่ 2 ได้รับอิทธิพลมาจากลักษณะของการกระจายของคะแนนสอบ (score distribution) ของกลุ่มที่นำมาเปรียบเทียบด้วย ดังนั้นข้อทดสอบจะถือว่ามีลักษณะลำเอียงหรือไม่ ยังขึ้นอยู่กับลักษณะการกระจายของคะแนนสอบด้วย ซึ่งเป็นปัจจัยที่อยู่นอกเหนือคำนิยามของความลำเอียงของวิธีนี้ (Ironson, 1982: 136)
 - ค. มีโอกาสที่เป็นไปได้ที่ข้อทดสอบมีความลำเอียงแต่วิธีการวิเคราะห์วิธีที่ 3 ไม่อาจค้นพบได้ เนื่องจากข้อทดสอบที่ค่าพารามิเตอร์ต่างกันอาจมีโค้งลักษณะของข้อทดสอบ (ICC) เหมือนกันได้ (Ironson, 1982: 49)

ดังนั้น จากข้อบกพร่องทั้ง 3 ข้างต้นนี้ จึงทำให้กล่าวได้ว่าจำนวนข้อทดสอบที่ลำเอียงนั้นอาจมีบางข้อที่ไม่ใช่ข้อที่ลำเอียง และอาจมีบางข้อที่วิธีวิเคราะห์ทั้ง 3 ไม่อาจค้นพบได้ และจากการที่จำนวนข้อทดสอบที่ลำเอียงที่เพิ่มขึ้นหรือลดลงอย่างไม่มีระบบ รวมทั้งการค้นหาค่าความลำเอียงของแต่ละวิธี มุ่งวิเคราะห์ลักษณะของ

ข้อทดสอบคนละด้าน จึงอาจเป็นเหตุให้ จำนวนข้อทดสอบที่ลำเอียงของแต่ละวิธี มีความสัมพันธ์กันอย่างไม่มีนัยสำคัญ ผลการวิจัยครั้งนี้ขัดแย้งกับผลการวิจัยของ Rudner และคณะ (Rudner and Others, 1980: 8) ซึ่งศึกษาเปรียบเทียบ ผลของการวิเคราะห์หาข้อทดสอบที่ลำเอียง 7 วิธี โดยอาศัยข้อมูลที่สร้างขึ้นจากวิธี Monte Carlo และพบว่าค่าสัมประสิทธิ์สหสัมพันธ์ของจำนวนข้อทดสอบที่พบจาก 3 วิธีที่ผู้วิจัยศึกษาครั้งนี้มีค่าในระดับปานกลางคือ r_{xy} ระหว่าง 0.60-0.66 ซึ่ง Rudner และคณะกล่าวว่า มีค่าสูงกว่าผลการวิจัยที่พบโดย Ironson (1979) และ Rudner และ Coney (1978) ซึ่งศึกษาจากข้อมูลจริงที่มีขนาด น้อยกว่า (อ้างจาก Rudner และคณะ, 1980: 9) แต่ข้อค้นพบที่ขัดแย้งกันนี้อาจ เป็นไปได้ เพราะการศึกษาค้นคว้าครั้งนี้ผู้วิจัยอาศัยข้อมูลเชิงประจักษ์ (empirical data) จึงมีจำนวนการกระทำซ้ำ (repetition) ในการวิเคราะห์น้อยครั้ง กว่าการศึกษาที่ Rudner และคณะจัดทำขึ้นกับข้อมูลจำลอง (simulated data) ซึ่งกระทำซ้ำได้จำนวนมากมายและภาสได้หลายเงื่อนไข

ข้อเสนอแนะ

- ก. สำหรับคณะกรรมการสร้างแบบทดสอบภาษาอังกฤษเข้ามหาวิทยาลัยและผู้ที่เกี่ยวข้อง
 1. ควรทำการศึกษาและวิเคราะห์ข้อทดสอบที่ผลการวิจัยครั้งนี้ พบว่ามีความลำเอียงในลักษณะต่าง ๆ ในเชิงของการวิเคราะห์เนื้อหาและตรรก (Judgmental Analysis) เพื่อค้นหาว่าเหตุใดข้อทดสอบเหล่านั้นจึงมีความลำเอียงต่อเพศ หรือต่อภาคภูมิศาสตร์ของผู้สอบ ทั้งนี้เพื่อให้ได้แนวทาง ในการเลือกเนื้อหาและการสร้างข้อทดสอบภาษาอังกฤษเข้ามหาวิทยาลัยต่อไป เพื่อให้ข้อทดสอบมีความเป็นธรรมแก่ผู้สอบให้ได้มากที่สุด เพราะ "การตัดสินใจให้ผู้สมควรสอบได้ตกและตัดสินใจให้ผู้สมควรสอบตกได้" เป็น ความผิดพลาดที่ร้ายแรงที่คณะกรรมการสร้างข้อทดสอบควรตระหนักไว้ เสมอ และควรพยายามทุกวิถีทางให้เกิดเหตุการณ์ดังกล่าวให้น้อยที่สุด เพราะการตัดสินใจที่ผิดพลาดดังกล่าวมีผลกระทบต่ออนาคตของบุคคลทั้ง ผู้ที่สอบได้และสอบตกและคุณภาพของประชากรของประเทศเป็นอย่างมาก
 2. ควรศึกษาข้อเสนอแนะในการสร้างและพัฒนาข้อทดสอบเกี่ยวกับการแก้ไข ปัญหาความลำเอียงของข้อทดสอบ (Burrill, 1982: 175-177) เช่น
- ไม่ควรเชื่อว่าข้อทดสอบมีความลำเอียงจากค่าสถิติเท่านั้น

- ถ้าพัฒนาแบบทดสอบ โดยอาศัยทฤษฎีการตอบสนองของข้อทดสอบ (Item Response Theory) ควรวิเคราะห์หาข้อทดสอบที่ลำเอียง โดยวิธี Three-parameter Logistic Method
- ถ้าพัฒนาแบบทดสอบ โดยวิธีประเพณีนิยมควรวิเคราะห์หาข้อทดสอบที่ลำเอียง โดยวิธี Chi-square Method หรือ Delta-Plot Method

ข. สำหรับนักวิจัย

1. ควรทำการศึกษาและวิเคราะห์ข้อทดสอบที่ผลการวิจัยครั้งนี้เห็นว่ามีความลำเอียงในเชิงของการวิเคราะห์เนื้อหาและตรรกว่าเป็นเพราะเหตุใด สอดคล้องกับความเป็นจริงมากน้อยเพียงใด และวิธีวิเคราะห์ใดที่ให้ผลสอดคล้องกับความเป็นจริงมากที่สุด
2. ควรทำการศึกษาในทำนองเดียวกันกับการวิจัยครั้งนี้ โดยใช้ข้อทดสอบภาษาอังกฤษ เข้ามหาวิทยาลัย ชุด กษ. และ กษค. ปี พ.ศ. 2534 เป็นต้นมา เพื่อตรวจสอบคุณภาพของแบบทดสอบวิธีหนึ่ง และเพื่อศึกษาว่าผลการวิจัยสอดคล้องหรือขัดแย้งกับผลการวิจัยนี้หรือไม่ โดยเฉพาะอย่างยิ่ง เรื่องความสัมพันธ์ของจำนวนข้อทดสอบที่ลำเอียง ซึ่งได้จากการวิเคราะห์แต่ละวิธี เนื่องจากเป็นประเด็นที่น่าสงสัยอยู่

ค. สำหรับผู้สร้างและพัฒนาแบบทดสอบทั่วไป

1. ควรศึกษาข้อเสนอแนะในการสร้างและพัฒนาข้อทดสอบเกี่ยวกับการแก้ไข ปัญหาความลำเอียงของข้อทดสอบ ดังที่ได้อ้างถึงมาแล้วในข้อ 2 ของ ข้อ ข
2. เนื่องจากมีผลการวิจัยจำนวนมากค้นพบและยืนยันว่า การวิเคราะห์หาความลำเอียง ที่ได้ผลดีที่สุด เพราะให้รายละเอียดได้มากที่สุดและสอดคล้องกับแนวคิดทางการทดสอบ แนวใหม่มากที่สุด คือ Three-parameter Logistic Method รองลงมาได้แก่ Chi-square Method ที่แบ่งผู้สอบออกเป็นกลุ่มย่อย 5 กลุ่ม (Subkoviak and Others, 1984: 56; Rudner and Others, 1980: 9; Ironson, 1979 อ้างจาก Burrill, 1982: 169) ดังนั้น หากเป็นไปได้ควรวิเคราะห์

หาความลำเอียงของข้อทดสอบ โดย Three-parameter Logistic Method แต่วิธีดังกล่าวนี้เสียค่าใช้จ่าย และเวลาในการคำนวณมากที่สุด รวมทั้งต้องอาศัยผู้สอบมากกว่า 1,000 คน และโปรแกรมคอมพิวเตอร์ด้วย ดังนั้น หากไม่สามารถกระทำได้ในทางปฏิบัติก็ควรใช้ Chi-square Method แทน หรือหากยังไม่สามารถกระทำได้ ก็ควรใช้ Delta-Plot Method ซึ่งเป็นวิธีที่ง่ายที่สุด และสะดวกที่สุดในทางปฏิบัติ

3. ข้อทดสอบที่ลำเอียงไม่จำเป็นที่จะต้องเป็นข้อทดสอบที่ "ไม่ดี" เสมอไป เพราะในบางสถานการณ์ ข้อทดสอบที่ลำเอียงอาจให้ข้อสารสนเทศที่ดีที่สุด (most information) หรือเป็นตัวทำนายที่ดีที่สุด (best predictor) เช่น ในการเลือกวิชาเรียน (Rudner and Others, 1984 b: 227) ดังนั้นในการพัฒนาข้อทดสอบควรคำนึงถึงเรื่องวัตถุประสงค์ของแบบทดสอบเป็นลำดับด้วยเสมอ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



บรรณานุกรม

- ไกรทอง อนันตกุล. "การศึกษาผลสัมฤทธิ์ในการเรียนโครงสร้างไวยากรณ์อังกฤษของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ในกรุงเทพมหานคร." วิทยานิพนธ์การศึกษามหาบัณฑิต. กทม.: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2525.
- คณะกรรมการการศึกษาแห่งชาติ. การวิจัยและประเมินคุณภาพโรงเรียนประถมศึกษา. กทม.: สำนักงานคณะกรรมการการศึกษาแห่งชาติ สำนักงานข้าราชการครู, 2531.
- เฉลิมชัย มิติภักษ์. "การศึกษาความสามารถในการฟังภาษาอังกฤษของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ในโรงเรียนมัธยมศึกษาที่ตั้งอยู่ในเขตอำเภอเมืองและโรงเรียนที่ตั้งอยู่นอกเขตอำเภอเมือง." วิทยานิพนธ์การศึกษามหาบัณฑิต. กทม.: บัณฑิตวิทยาลัย มหาวิทยาลัยเกษตรศาสตร์, 2522.
- ธีชชัย เผ่าพงษ์. "การวิเคราะห์ความลำเอียงของข้อสอบมาตรฐานวัดความถนัดด้านคณิตศาสตร์ และภาษาของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 3." วิทยานิพนธ์การศึกษามหาบัณฑิต. ขอนแก่น: บัณฑิตวิทยาลัย มหาวิทยาลัยขอนแก่น, 2526.
- ประทุม เจริญผล. "องค์ประกอบบางประการที่มีความสัมพันธ์กับผลสัมฤทธิ์ทางการเรียนภาษาอังกฤษของนักศึกษาผู้ใหญ่สายสามัญ ระดับที่ 5 ในเขตการศึกษา 5." วิทยานิพนธ์การศึกษามหาบัณฑิต. นครปฐม: บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร, 2526.
- ฝ่ายสารสนเทศ 4. "สถิติการศึกษา". วารสารการศึกษาแห่งชาติ. 2 (ธันวาคม, 2529 - มกราคม 2530): 45-50.
- ทัศนีย์ พิรมนตรี. "การวิเคราะห์ความลำเอียงของแบบสอบวิชาคณิตศาสตร์ โครงการตรวจสอบคุณภาพการศึกษา ชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526." วิทยานิพนธ์ครุศาสตร์มหาบัณฑิต. กทม.: บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2530.
- สุพงษ์ สุกมลสันต์. "คุณภาพ คุณธรรม และแนวโน้มของข้อทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย" ใน การวิเคราะห์ข้อทดสอบเข้ามหาวิทยาลัยวิชาภาษาอังกฤษ. Priya Unaratana (ed.) กทม.: สมาคม Thai/TESOL Occasional papers No. 1990/1, 1990.
- _____. "ความสอดคล้องของข้อทดสอบคัดเลือกเข้ามหาวิทยาลัยกับเกณฑ์ในทฤษฎีการวัดผล". ใน ความสอดคล้องระหว่างข้อสอบเข้ามหาวิทยาลัยกับหลักสูตรภาษาอังกฤษ และแนวโน้มในการเรียนการสอนภาษาอังกฤษ เพื่อการสื่อสารและการสร้างคลังข้อสอบ. กทม.: Thai/TESOL's Occasional Papers No. 1984/1, 1984.
- _____. "ความสอดคล้องของข้อทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย ชุด กข. และ กขค. ปี 2528 กับเกณฑ์ในทฤษฎีการวัดผล". บทความเสนอในการประชุมใหญ่ของสมาคม Thai/TESOL เรื่อง การวิเคราะห์ข้อทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย, 2528.

สุรศักดิ์ อมรรัตนศักดิ์. "การศึกษาเปรียบเทียบผลของวิธีวิเคราะห์ความลำเอียงของข้อสอบ
ที่แตกต่างกัน 4 วิธี." วิทยานินนธ์ครุศาสตร์ศรีบัณฑิต. กทม.: บัณฑิตวิทยาลัย จุฬาลงกรณ์-
มหาวิทยาลัย, 2531.

อำไพ อัมมัตต์ศฤงคาร. "สัมฤทธิ์ผลทางการเรียนวิชาภาษาอังกฤษชั้นมัธยมศึกษาปีที่ 3
โรงเรียนมหาวชิราวุธ." วิทยานินนธ์ครุศาสตร์มหาบัณฑิต. กทม.: บัณฑิตวิทยาลัย
จุฬาลงกรณ์มหาวิทยาลัย, 2516.



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

- Angoff, W.H. "Use of Difficulty and Discrimination Indices for Detecting Item Bias." In Handbook of Methods for Detecting Test Bias, R.A. Berk (ed.) Maryland: The Johns Hopkins University Press, 1982.
- _____ and Sharon, A.T. "The Evaluation of differences in Test Performance of Two or More Groups". Educational and Psychological Measurement. 34 (1974): 807-816.
- Baker, F.B. "Advances in Item Analysis." Review of Educational Research. 17 (Winter 1977): 151-178.
- Berk, R.A. Handbook of Methods for Detecting Test Bias. Maryland: The Johns Hopkins University Press, 1982.
- Burstall, C. "Factors Affecting Foreign Language Learning: A Consideration of Some Recent Research Findings." In Language Teaching and Linguistics, V. Kinsella (ed.) London: Oxford University Press, 1978.
- Burrill, L.E. "Comparative Studies of Item Bias Methods." In Handbook of Methods for Detecting Test Bias, R.A. Berk (ed.) Maryland: The Johns Hopkins University Press, 1982.
- Bradford, A. "A farang's Problems with the University Entrance Examinations in English." ใน การวิเคราะห์ข้อทดสอบเข้ามหาวิทยาลัยวิชาภาษาอังกฤษ. Priya Unaratana (ed.) กทม. สมาคม Thai/TESOL Occasional Papers No. 1991/1, 1990.
- Carrell, P.L. "Schema Theory and ESL Reading: Classroom Implications and Applications." The Modern Language Journal. 68 (1984): 332-342.
- Carton, S.T. and Marco, G.L. "Methods Used by Test Publishers." In Handbook of Methods for Detecting Test Bias, R.A. Berk (ed.) Maryland: The Johns Hopkins University Press, 1982.
- Cleary, T.A. and Hilton, T.L. "An Investigation of Item Bias." Educational and Psychological Measurement. 28 (1968): 61-75.
- Hambleton, R.K. and Cook, L.L. "Latent Trait Models and Their Use in the Analysis of Educational Test Data." Journal of Educational Measurement. 14 (Summer 1977): 75-95.

- Intasuwan, P. "A Comparison of Three Approaches for Detecting Item Bias in Cross-national Testing." Dissertation Abstracts International, 1979, 40, 2613A.
- Ironson, G.H. "Use of Chi-square and Latent Trait Approaches for Detecting Items Bias." In Handbook of Methods for Detecting Test Bias, R.A. Berk (ed.) Maryland: The Johns Hopkins University Press, 1982.
- Ironson, G.H. and Subkoviak, M.J. "A Comparison of Several Methods of Assessing item Bias." Journal of Educational Measurement 16 (1979): 209-225.
- Lakana, S. "Application of Analysis of Variance Approach and Item Characteristic Approach for Assessing Item Bias in the ITBS Form 7." Dissertation Abstracts International, 1979, 40, 2615A.
- Lord, F.M. Applications of Item Response Theory to Practical Testing Problems. New Jersey: Lawrence Erlbaum, 1980.
- _____ and Novick, M.R. Statistical Theories of Mental Test Scores. California: Addison-Wesley Publishing Co., 1968.
- Mandler, J.M. "Categorical and Schematic Organization in Memory." In Memory Organization and Structure, C.R. Puff (ed.). New York: Academic Press, 1979.
- Osterlind, S.J. Test Item Bias, California: SAGE Publications, Inc. 1983.
- Rudner, L.M. and Others. "A Monte Carlo Comparison of Seven Biased Item Detection Techniques." Journal of Educational Measurement. 17 (Spring 1980): 1-10.
- _____. "Biased Item Detection Techniques." Journal of Educational Statistics. 5 (Fall 1980 b): 213-233.
- Shepard, L.A. "Definitions of Bias." In Handbook of Methods for Detecting Test Bias, R.A. Berk (ed.) Maryland: The Johns Hopkins University Press, 1982.
- SPSSX User's Guide. New York: McGraw-Hill Book Company, 1983.

- Subkoviack, M.J. and Others. "Empirical Comparison of Selected Item Bias Detection Procedures with Bias Manipulation." Journal of Educational Measurement. 21 (Spring 1984): 49-58.
- Warm, T.A. A Primer of Item Response Theory. Oklahoma: US Coast Guard Institute, 1978.
- Wingersky, M.s. A User's Guide to LOGIST 5. Princeton New Jersey: Educational Testing Service, 1982.
- Yamane, T. Statistics: An Introductory Analysis. New York: Harper and Row, 1967.



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



ภาคผนวก ก. ข. และ ค.
เป็นค่าสถิติที่เกี่ยวข้องกับผลการ
วิเคราะห์ที่ออกสอบแต่ละข้อและ
แต่ละวิชาซึ่งมีรายละเอียดมากและ
ปริมาณมาก ดังนั้นจึงจึงไม่ได้
รวมไว้ในรายงานนี้ ผู้สนใจอาจ
ศึกษาจากฉบับสมบูรณ์ได้ที่ผู้วิจัย

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ง

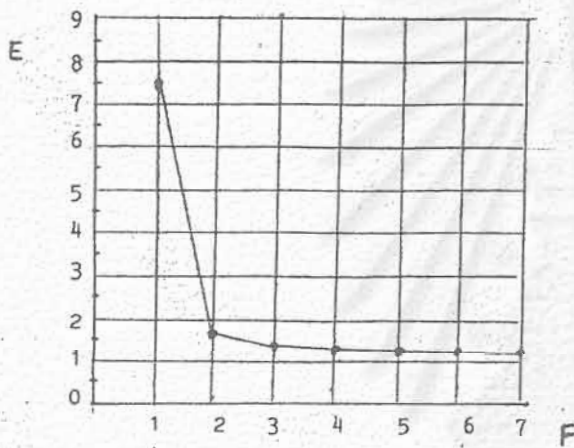
สถิติ โปรแกรม และข้อมูลบางส่วนที่เกี่ยวข้อง



สถาบันวิทย์บริการ
จุฬาลงกรณ์มหาวิทยาลัย

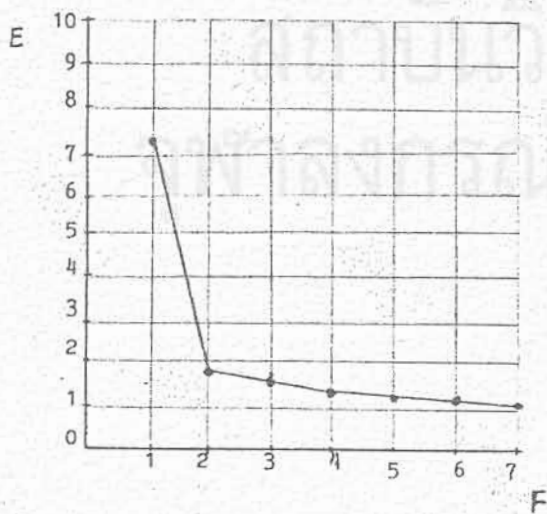
การทดสอบความมีมิติเดียว (Unidimensionality ของแบบทดสอบ)

กราฟที่ 1: แบบทดสอบภาษาอังกฤษชุด กข ปี พ.ศ. 2531



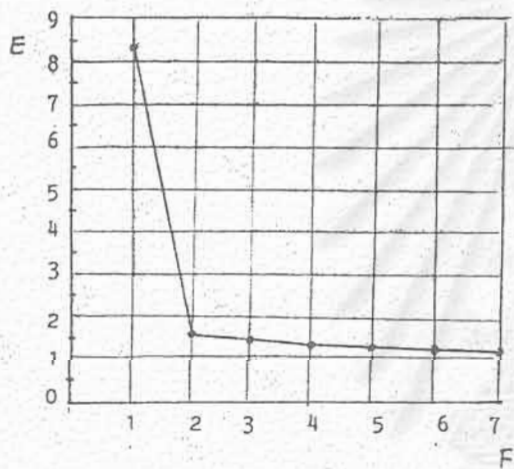
Factor ที่	Eigenvalue
1	7.5654
2	1.8524
3	1.6552
4	1.3361
5	1.2660
6	1.2544
7	1.2267

กราฟที่ 2: แบบทดสอบภาษาอังกฤษชุด กข. ปี พ.ศ. 2532



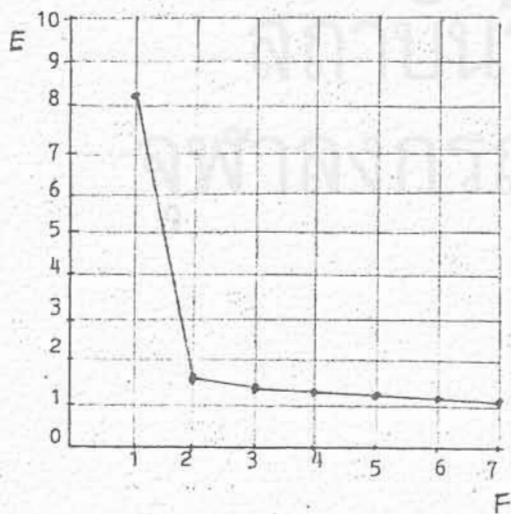
Factor ที่	Eigenvalue
1	7.2693
2	1.9984
3	1.9615
4	1.4164
5	1.3585
6	1.2503
7	1.1894

กราฟที่ 3: แบบทดสอบภาษาอังกฤษชุด กข ปี พ.ศ. 2533



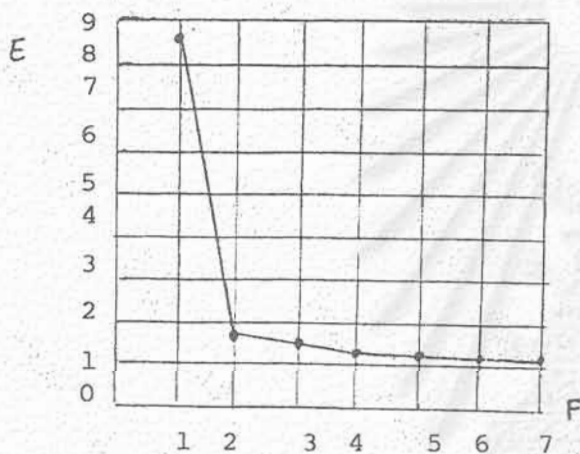
Factor ที่	Eigenvalue
1	8.4855
2	1.6043
3	1.5621
4	1.4313
5	1.4054
6	1.2644
7	1.2118

กราฟที่ 4: แบบทดสอบภาษาอังกฤษชุด กขค. ปี พ.ศ. 2531



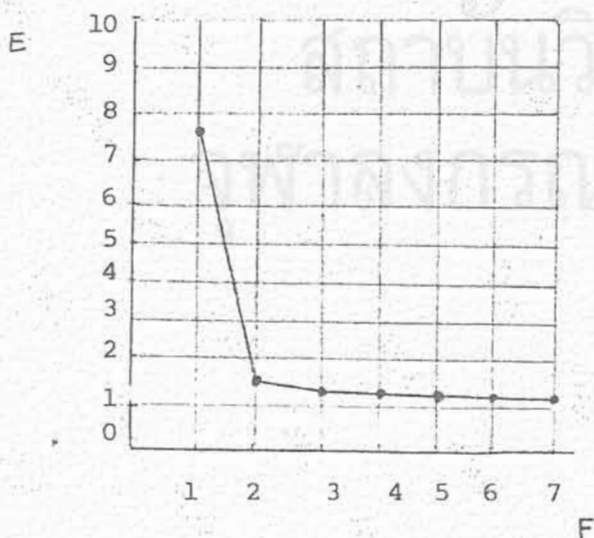
Factor ที่	Eigenvalue
1	8.1797
2	1.6427
3	1.5243
4	1.4385
5	1.2468
6	1.2213
7	1.1994

กราฟที่ 5: แบบทดสอบภาษาอังกฤษชุด กขค. ปี พ.ศ. 2532



Factor ที่	Eigenvalue
1	8.7475
2	1.9652
3	1.8573
4	1.3581
5	1.2885
6	1.2211
7	1.2114

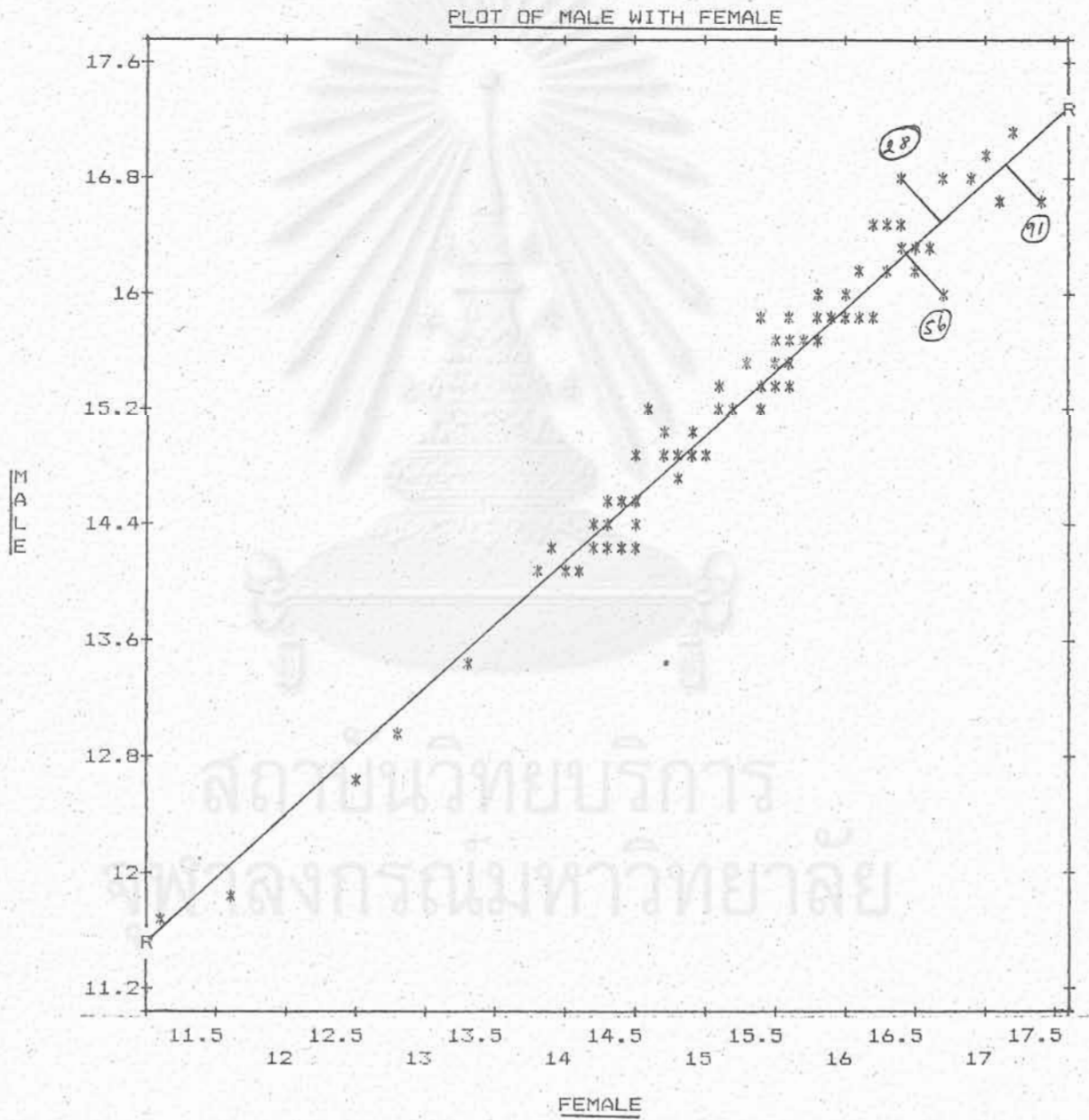
กราฟที่ 6: แบบทดสอบภาษาอังกฤษชุด กขค. ปี พ.ศ. 2533



Factor ที่	Eigenvalue
1	7.7428
2	1.6218
3	1.4527
4	1.3803
5	1.3041
6	1.2684
7	1.2132

ตัวอย่างลักษณะของกราฟและระยะห่างตั้งฉากจากแกนหลัก (d_1)

กรณีของแบบทดสอบภาษาอังกฤษชุด กข. ปี 2531



ตัวอย่างการคำนวณหาความลำเอียงของข้อทดสอบ โดยวิธี
(ข้อที่ 4 ชุด กข. ปี 2531)

Chi-square Method

กลุ่มคะแนนต่ำ

	ชาย	หญิง	รวม
ผิด	705	597	1302
	5401	45.9	50.0
	52.4	47.4	
ถูก	641	662	1303
	49.2	50.8	50.0
	47.6	52.6	
รวม	1346	1259	26.5
	51.7	48.2	100.0

$$\chi^2 = 6.2017$$

กลุ่มคะแนนสูง

	ชาย	หญิง	รวม
ผิด	228	268	496
	46.0	54.0	40.7
	39.1	42.1	
ถูก	355	368	723
	49.1	50.9	59.3
	60.9	57.9	
รวม	583	636	1219
	47.8	52.2	100.0

$$\chi^2 = 1.0352$$

กลุ่มคะแนนปานกลาง

	ชาย	หญิง	รวม
ผิด	601	611	1212
	49.6	50.4	55.7
	56.1	55.3	
ถูก	470	494	964
	48.8	51.2	44.3
	43.9	44.7	
รวม	1071	1105	2176
	49.2	50.8	100.0

$$\chi^2 = 0.1173$$

การตัดสินใจ

$$\alpha = 0.05$$

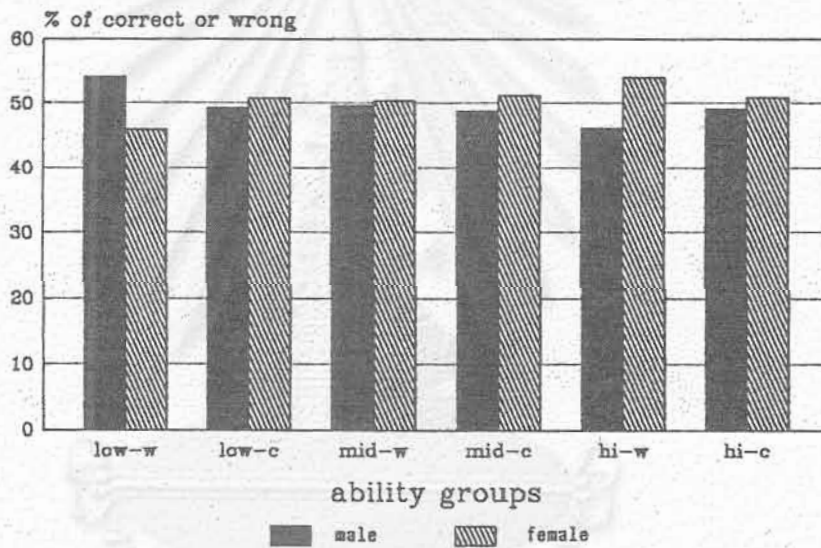
$$\text{เกณฑ์การตัดสินใจ} = \frac{\alpha}{.k} = \frac{0.05}{3} = 0.0167$$

$$\text{และ } \chi^2_{1,0.0167} = 5.7246$$

ดังนั้น ข้อทดสอบข้อที่ 4 มีความลำเอียงต่อผู้สอบ
กลุ่มคะแนนต่ำ โดยมีความลำเอียงต่อผู้สมัครสอบ
เพศชาย (เนื่องจากอัตราส่วนของผู้ที่ตอบถูกเป็น
ผู้สมัครสอบเพศชายน้อยกว่าเพศหญิง ทั้ง ๆ ที่
เป็นผู้ที่มีคะแนนรวมอยู่ในกลุ่มคะแนนรวมเดียวกัน)

ตัวอย่างกราฟอัตราส่วนการตอบถูกหรือผิดของข้อทดสอบ เมื่อวิเคราะห์ด้วย Chi-square Method
(ข้อที่ 4 ของภาษาอังกฤษชุด กข. ปี พ.ศ. 2531)

ITEM BIASES AGAINST SEXES chi-square method



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตัวอย่างการคำนวณหาความลำเอียงของข้อทดสอบ โดยวิธี χ^2 (ข้อที่ 5 ชุด กข. ปี 2531)

กลุ่มคะแนนต่ำ

	ชาย	หญิง	รวม
ผิด	1197	1133	2330
	51.4	48.6	89.4
	88.9	90.0	
ถูก	149	126	275
	54.2	45.8	10.6
	11.1	10.0	
รวม	1346	1259	2605
	51.7	48.3	100.0

$$\chi^2 = 0.6684$$

กลุ่มคะแนนสูง

	ชาย	หญิง	รวม
ผิด	354	405	759
	46.6	53.4	62.3
	60.7	63.7	
ถูก	229	231	460
	49.8	50.2	37.7
	39.3	36.3	
รวม	583	636	1219
	47.8	52.2	100.0

$$\chi^2 = 1.0109$$

กลุ่มคะแนนปานกลาง

	ชาย	หญิง	รวม
ผิด	844	886	1730
	48.8	51.2	79.5
	78.8	80.2	
ถูก	227	219	446
	50.9	49.1	20.5
	21.2	19.8	
รวม	1071	1105	2176
	49.2	50.8	100.0

$$\chi^2 = 0.5504$$

การตัดสินใจ

$$\chi^2_{1,0.0167} = 5.7246$$

ดังนั้น ข้อทดสอบข้อนี้จึงไม่มีความลำเอียงต่อผู้สมัครสอบเพศชายหรือเพศหญิงในทุก ๆ กลุ่มระดับคะแนน (เนื่องจากอัตราส่วนของผู้สมัครสอบเพศชายหรือเพศหญิง ในการตอบถูกหรือผิดใกล้เคียงกันและคล้อยตามกันในทุก ๆ กลุ่มระดับคะแนน)

การคำนวณหาพื้นที่ของความแตกต่างระหว่าง ICC

1. สูตรทั่วไปของโอกาสในการตอบข้อทดสอบข้อที่ i ถูก (เมื่อใช้ 3 parameters)

$$P(\theta) = C + (1-C) \frac{\text{Exp}^{D_a(\theta-b)}}{1 + \text{Exp}^{D_a(\theta-b)}} \quad \text{เมื่อ } D = 1.7$$

2. ในกรณีที่กำหนดให้ θ มีค่า ± 5 ICC ของข้อทดสอบข้อที่ i คือ

$$\int_{-5}^5 P(\theta) \Delta\theta$$

3. สมการที่ 1 และ 2 อาจเขียนได้อีกในรูปแบบหนึ่งดังนี้

$$\int_{-5}^5 P(\theta) \Delta\theta = (C \times 10) + \frac{1-C}{D_a} \left[\ln \left[1 + e^{D_a(5-b)} \right] - \ln \left[1 + e^{D_a(-5-b)} \right] \right]$$

4. ความแตกต่างของพื้นที่ระหว่าง ICC 2 เส้นคือ

$$\phi_i = \int_{-5}^5 P_1(\theta) \Delta\theta - \int_{-5}^5 P_2(\theta) \Delta\theta$$

5. การคำนวณหาพื้นที่ระหว่าง ICC 2 เส้นของผู้สอบ 2 กลุ่ม อาจทำได้โดยใช้โปรแกรมคอมพิวเตอร์ และ ICC ที่ได้อาจมีลักษณะแตกต่างกันดังตัวอย่างต่อไปนี้ (การคำนวณและการ plot กราฟในตัวอย่างนี้ใช้โปรแกรมสำเร็จรูปชื่อ MathCad)

จุฬาลงกรณ์มหาวิทยาลัย

ตัวอย่างผลการวิเคราะห์หาความลำเอียงด้วย Three-parameter Logistic Method

a := 0.5413

b := -0.1482

c := 0.0767

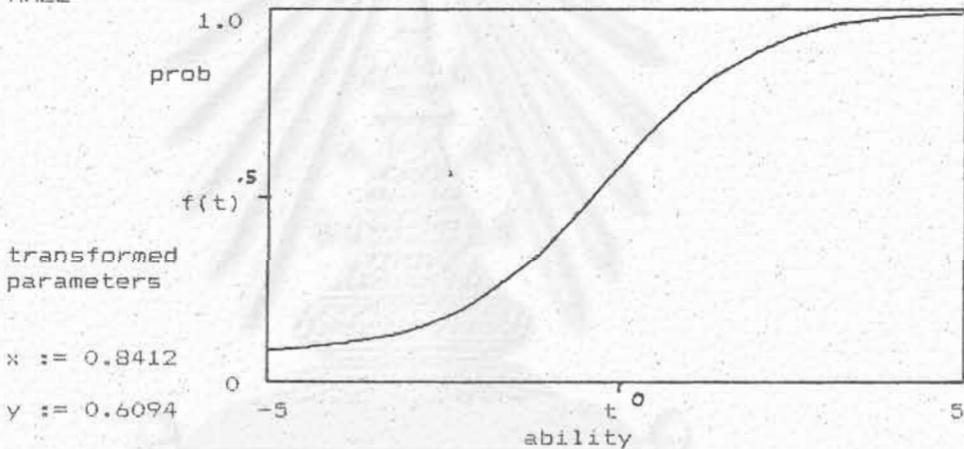
t := -5, -4.95 .. 5

d := 1.7

2531-ENGLISH EXAM FORM AB, ITEM NO. 41
ITEM BIASES AGAINST SEXES

$$f(t) := c + (1 - c) \frac{\exp(d \cdot a \cdot (t - b))}{1 + \exp(d \cdot a \cdot (t - b))} \quad \int_{-5}^5 f(t) dt = 5.518$$

MALE



x := 0.8412

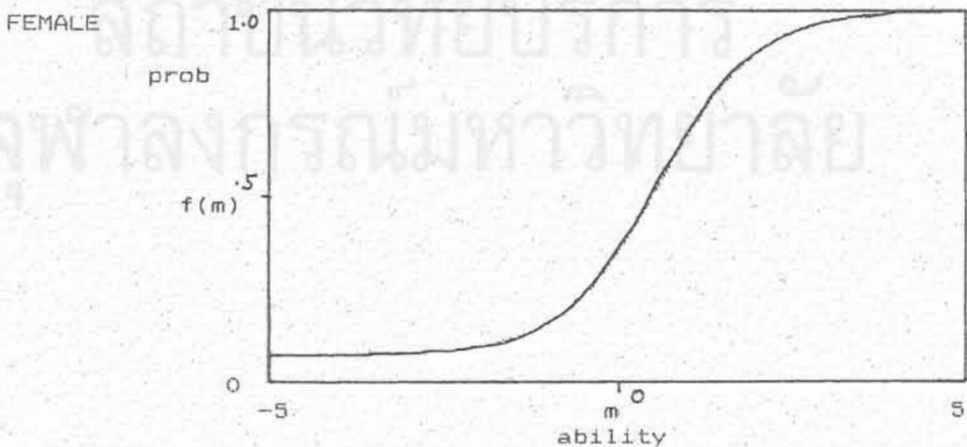
y := 0.6094

z := 0.0758

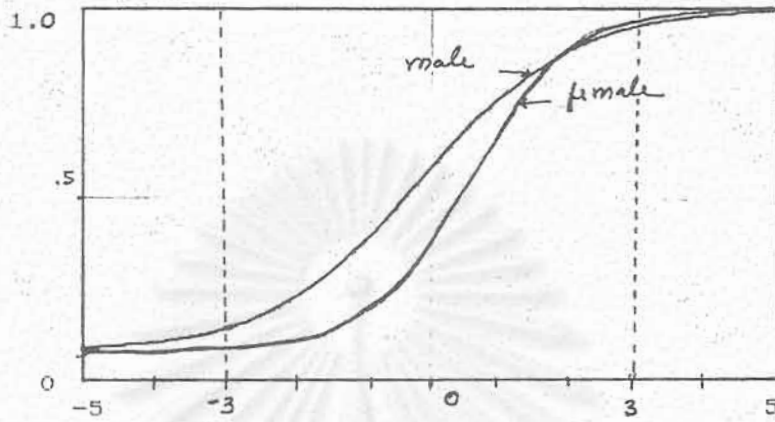
m := -5, -4.95 .. 5

$$f(m) := z + (1 - z) \frac{\exp(d \cdot x \cdot (m - y))}{1 + \exp(d \cdot x \cdot (m - y))} \quad \int_{-5}^5 f(m) dm = 4.817$$

FEMALE



ภาพซ้อน



ภาพขยายระหว่าง $\theta = \pm 3$

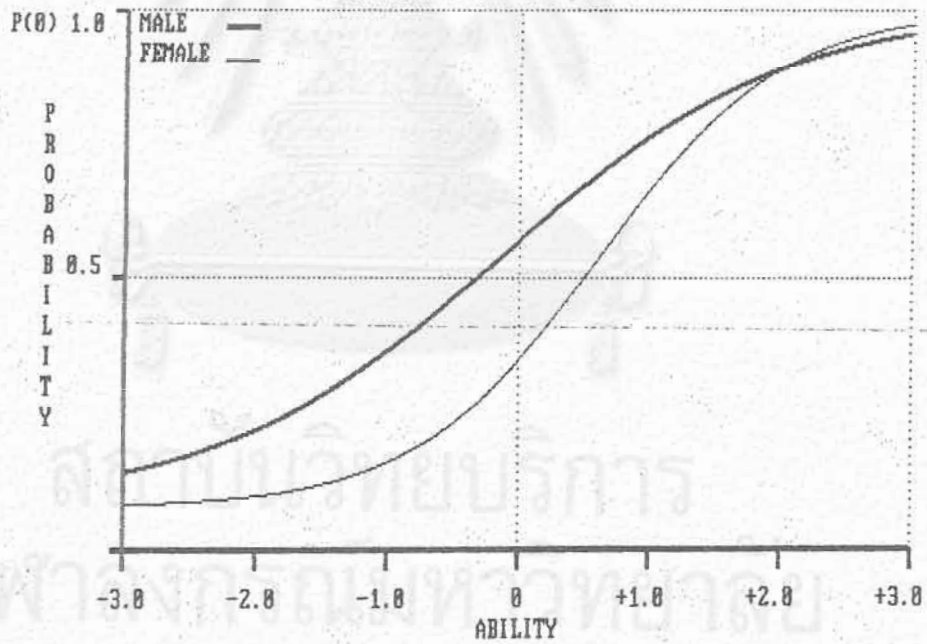


Figure 140

DETAIL OF BIAS

I_NO	FROM ABILIT	TO ABILITY	AGAINST	INDEX
1	-3.000000000E+00	1.979999999E+00	female	6.834703969E-01
2	1.979999999E+00	2.999999999E+00	male	1.4187977449E-02

TOTAL INDEX 6.9765837436E-01

a := 0.4107

b := 3.1177

c := 0.2882

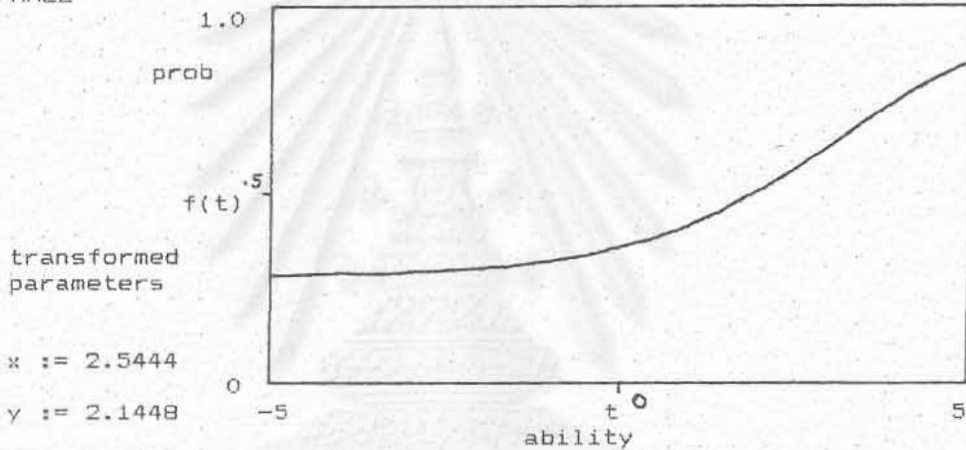
t := -5, -4.95 .. 5

d := 1.7

2531-ENGLISH EXAM FORM AB, ITEM NO. 74
ITEM BIASES AGAINST SEXES

$$f(t) := c + (1 - c) \frac{\exp(d \cdot a \cdot (t - b))}{1 + \exp(d \cdot a \cdot (t - b))} \quad \int_{-5}^5 f(t) dt = 4.461$$

MALE



x := 2.5444

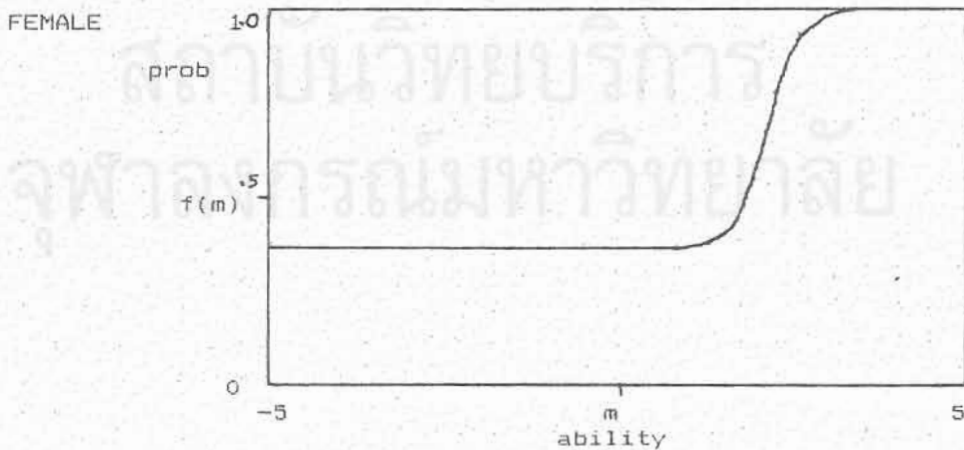
y := 2.1448

z := 0.3617

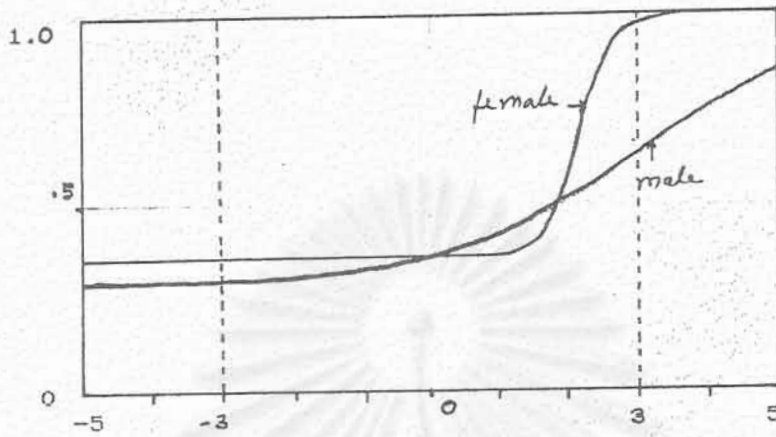
m := -5, -4.95 .. 5

$$f(m) := z + (1 - z) \frac{\exp(d \cdot x \cdot (m - y))}{1 + \exp(d \cdot x \cdot (m - y))} \quad \int_{-5}^5 f(m) dm = 5.439$$

FEMALE



ภาพซ้อน



ภาพขยายระหว่าง $\theta = \pm 3$

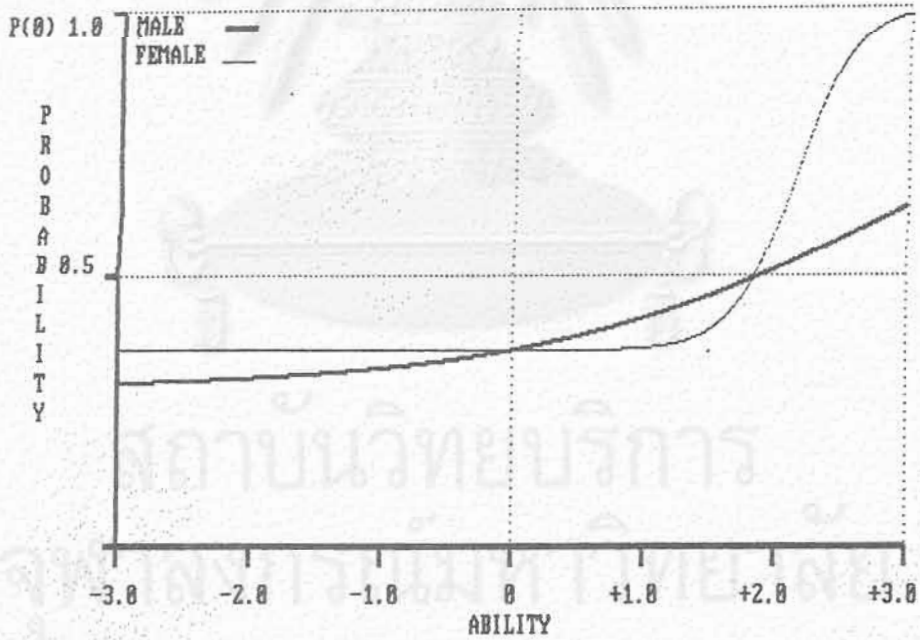


Figure 174

DETAIL OF BIAS

I_NO	FROM ABILIT	TO ABILITY	AGAINST	INDEX
1	-3.0000000000E+00	2.9999999638E-02	male	1.2514080377E-01
2	2.9999999638E-02	1.8399999995E+00	female	7.0421560848E-02
3	1.8399999995E+00	2.999999993E+00	male	2.8426926622E-01
TOTAL INDEX				4.7983163084E-01

โปรแกรมที่เขียนขึ้นใช้เฉพาะงาน

โปรแกรมที่ 1: สำหรับ Delta-Plot Method

```
CLS
*****
* THIS IS THE PROGRAM USED TO FIND Di OF AN ITEM WHEN ANALYZED *
* BY TID/DPM METHOD. IT IS WRITTEN BY SUPHAT SUKAMOLSON USING *
* QUICK-BASIC VERSION 4.5 IN 1991. *
* DELTAs OF A BASE GROUP ARE Ps AND THOSE OF A TARGET GROUP *
* ARE PDs. *
*****
DIM P(100), PD(100)
X = 0
Y = 0
XX = 0
YY = 0
XY = 0
OPEN "I", #1, "A:DATA6.TXT"
FOR I = 1 TO 100
INPUT #1, P(I), PD(I), N3, N4, N5, N6, N7, N8, N9, N10, N11, N12
X = X + P(I)
XX = XX + P(I) ^ 2
Y = Y + PD(I)
YY = YY + PD(I) ^ 2
Z = P(I) * PD(I)
XY = XY + Z
NEXT I
CLOSE #1
'FIND MEANS AND SDs OF THE 2 GROUPS AND ALSO Rxy
M1 = (X / 100)
M2 = (Y / 100)
S1 = SQR(((100 * XX) - (X ^ 2)) / 100 ^ 2)
S2 = SQR(((100 * YY) - (Y ^ 2)) / 100 ^ 2)
R1 = (100 * XY) - (X * Y)
RX = ((100 * XX) - X ^ 2)
RY = ((100 * YY) - Y ^ 2)
R2 = SQR(RX * RY)
R = R1 / R2
'FIND A VALUES (INTERCEPTS) OF A LINEAR EQUATION
'A VALUE IN CASE OF POSITIVE
AA = (S2 ^ 2 - S1 ^ 2) + SQR(((S2 ^ 2 - S1 ^ 2) ^ 2) +
(4 * R ^ 2 * S1 ^ 2 * S2 ^ 2))
AB = 2 * R * S1 * S2
A1 = AA / AB
'A VALUE IN CASE OF NEGATIVE
AC = (S2 ^ 2 - S1 ^ 2) - SQR(((S2 ^ 2 - S1 ^ 2) ^ 2) +
(4 * R ^ 2 * S1 ^ 2 * S2 ^ 2))
A2 = AC / AB
```

```

'FIND B VALUES (SLOPES) OF A LINEAR EQUATION
'B VALUE IN CASE OF POSITIVE A
B1 = M2 - A1 * M1
'B VALUE IN CASE OF NEGATIVE A
B2 = M2 - A2 * M1
'FIND Di (distance from the main axis)
'IN CASE OF POSITIVE A
FOR I = 1 TO 100
D1 = (A1 * P(I) + B1 - PD(I)) / SQR(A1 ^ 2 + 1)
'IN CASE OF NEGATIVE A
D2 = (A2 * P(I) + B2 - PD(I)) / SQR(A2 ^ 2 + 1)
PRINT USING "###.####"; I; D1; D2
LPRINT USING "###.####"; I; D1; D2
NEXT I
'FIND SD OF Di AND BIAS CRITERION AT P=0.05
SDE = SQR(1 - R)
CRIT = 1.964 * SDE
PRINT USING "#.####"; SDE; CRIT
LPRINT USING "#.####"; SDE; CRIT
LPRINT M1, S1, M2, S2, R, B1, B2, A1, A2
END

```

สถาบันวิทยบริการ

จุฬาลงกรณ์มหาวิทยาลัย

โปรแกรมที่ 2: สำหรับ Three-parameter Logistic Method

การคำนวณหาพื้นที่ของความแตกต่างระหว่าง ICC

```
CLS
*****
* THIS IS THE PROGRAM USED TO TRANSFORM PARAMETERS OF A TARGET GROUP *
* AND FIND THE AREA BETWEEN 2 ICCs OF A 3-PARAMETER LOGISTIC MODEL *
* WHEN THE DAs ARE BETWEEN -5 AND +5. IT'S WRITTEN BY SUPHAT SUKAMOLSON *
* USING QUICK-BASIC VERSION 4.5 IN 1991 *
*****
PARAMETERS OF A BASE GROUP ARE A, B & C
WHILE PARAMETERS OF A TARGET GROUP ARE AA, BB & CC
DIM A(100), B(100), C(100), AA(100), BB(100), CC(100), AT(100), BT(100)
OPEN "I", #1, "A:A01.TXT"
OPEN "I", #2, "A:A02.TXT"
SA = 0
XX = 0
SB = 0
YY = 0
'TO TRANSFORM PARAMETERS OF A TARGET GROUP
'FIND MEAN AND SD. OF THE BASE GROUP
FOR I = 1 TO 100
INPUT #1, A(I), B(I), C(I)
SA = SA + B(I)
XX = XX + B(I) ^ 2
IF EOF(1) THEN CLOSE #1
NEXT I
BX = (SA / 100)
S1 = (1 / 100) * SQR((100 * XX) - (SA ^ 2))
'FIND MEAN AND SD. OF THE TARGET GROUP
FOR I = 1 TO 100
INPUT #2, AA(I), BB(I), CC(I)
SB = SB + BB(I)
YY = YY + BB(I) ^ 2
IF EOF(2) THEN CLOSE #2
NEXT I
BY = (SB / 100)
S2 = (1 / 100) * SQR((100 * YY) - (SB ^ 2))
'FIND CONSTANT VALUES OF A & B FOR EQUATING
A1 = S1 / S2
B1 = BX - A1 * BY
'EQUATE AA & BB WITH A & B
FOR I = 1 TO 100
AT(I) = (1 / A1) * AA(I)
BT(I) = A1 * BB(I) + B1
PRINT USING "###.###"; I; A(I); B(I); C(I); AT(I); BT(I); CC(I)
LPRINT USING "###.###"; I; A(I); B(I); C(I); AT(I); BT(I); CC(I)
NEXT I
```

```

'TO FIND AREA BETWEEN 2 ICCs WHEN THEDAs = -5 TO +5
FOR I = 1 TO 100
K1 = EXP((1.7 * A(I)) * (5 - B(I)))
K2 = LOG(1 + K1)
K3 = EXP((1.7 * A(I)) * (-5 - B(I)))
K4 = LOG(1 + K3)
HM = (1 - C(I)) / (1.7 * A(I))
P1 = (C(I) * 10) + (HM * (K2 - K4))
L1 = EXP((1.7 * AT(I)) * (5 - BT(I)))
L2 = LOG(1 + L1)
L3 = EXP((1.7 * AT(I)) * (-5 - BT(I)))
L4 = LOG(1 + L3)
HN = (1 - CC(I)) / (1.7 * AT(I))
P2 = (CC(I) * 10) + (HN * (L2 - L4))
QQ = P1 - P2
PRINT USING "###.###"; I; QQ; P1; P2
LPRINT USING "###.###"; I; QQ; P1; P2
NEXT I
LPRINT BX, S1, BY, S2, A1, B1
END

```

สถาบันวิทยบริการ

จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ : ค่าสถิติพื้นฐานของแบบทดสอบ 6 ฉบับ เมื่อวิเคราะห์ในหลายลักษณะ

ปี	ฉบับ	ภาค/ เพศ	n	\bar{X}	S.D.	KR ₂₀	KR ₂₁	r _{tt}
2533	กข	C	1500	30.52	11.26	0.852	0.845	0.863
		E	1500	28.23	7.51	0.666	0.654	0.690
		NE	1500	27.64	7.26	0.645	0.634	0.677
		N	1500	29.65	9.43	0.787	0.778	0.813
		S	1500	28.63	7.93	0.699	0.688	0.714
		M	3000	28.78	8.97	0.766	0.758	0.781
		F	3000	28.98	8.70	0.805	0.741	0.772
	กขค	C	1500	28.08	8.52	0.744	0.734	0.764
		E	1354	25.93	6.35	0.548	0.536	0.566
		NE	484	225.81	6.26	0.537	0.524	0.498
		N	1500	26.14	6.98	0.626	0.617	0.638
		S	1500	26.34	6.53	0.570	0.558	0.613
		M	3000	26.16	6.76	0.600	0.590	0.616
		F	3000	26.99	7.54	0.676	0.666	0.692
2532	กข	C	1500	31.76	11.51	0.858	0.849	0.873
		E	1500	28.40	8.42	0.737	0.726	0.750
		NE	1500	28.35	8.83	0.761	0.752	0.776
		N	1500	30.23	10.32	0.824	0.814	0.830
		S	1500	28.70	8.65	0.750	0.739	0.751
		M	3000	29.43	9.85	0.807	0.798	0.817
		F	3000	30.04	10.09	0.816	0.806	0.831
	กขค	C	1500	29.73	10.97	0.846	0.838	0.851
		E	494	25.96	6.89	0.620	0.608	0.593
		NE	1176	26.36	7.21	0.648	0.638	0.654
		N	1186	27.19	7.86	0.702	0.692	0.689
		S	1392	26.10	6.97	0.626	0.615	0.641

ปี	ฉบับ	ภาค/ เขต	n	\bar{X}	S.D.	KR ₂₀	KR ₂₁	r _{tt}
		M	2874	26.82	7.95	0.710	0.701	0.716
		F	2874	27.62	8.98	0.773	0.764	0.776
2531	กข	C	1500	29.87	12.08	0.874	0.868	0.881
		E	1500	26.94	8.41	0.744	0.733	0.762
		NE	1500	26.62	8.49	0.750	0.741	0.763
		N	1500	28.32	10.51	0.836	0.828	0.834
		S	1500	27.30	8.78	0.763	0.754	0.768
		M	3000	27.84	10.03	0.819	0.812	0.826
		F	3000	28.15	10.22	0.826	0.818	0.832
	กขค	C	1500	34.32	10.96	0.836	0.825	0.830
		E	424	30.13	7.65	0.669	0.654	0.644
		NE	1170	29.48	8.17	0.714	0.701	0.731
		N	1052	31.15	8.79	0.749	0.735	0.753
		S	1334	29.07	7.91	0.697	0.683	0.696
		M	2740	30.62	9.10	0.767	0.756	0.766
		F	2740	31.54	9.44	0.782	0.770	0.783

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

