

รายงานการวิจัย

เรื่อง

การพัฒนาแบบทดสอบสมรรถภาพทางภาษาอังกฤษ
เพื่อวิชาชีพและการสื่อสารนานาชาติ
ของสถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย
และโปรแกรมการทดสอบให้ได้มาตรฐาน

**[A Standardization of CULI Tests of English
Proficiency for Occupational and International
Communication Purposes
and Their Testing Programs]**

รศ. ดร. สุวัฒน์ สุขมลันต์

สถาบันภาษา

จุฬาลงกรณ์มหาวิทยาลัย, 2549

งานวิจัยนี้ได้รับงบประมาณสนับสนุน
จากงบประมาณแผ่นดินประจำปี พ.ศ. 2546
ดังนั้น

ผู้วิจัยจึงใคร่ขอขอบคุณคณะกรรมการวิจัยของสถาบันภาษา
คณะกรรมการวิจัยของจุฬาลงกรณ์มหาวิทยาลัย
และคณะกรรมการวิจัยแห่งชาติ
ที่พิจารณาให้การสนับสนุนการวิจัยครั้งนี้

คำนำและกิตติกรรมประกาศ

รายงานวิจัยเรื่องนี้เป็นเรื่องที่เกิดจากโครงการวิจัยงบประมาณแผ่นดิน ปี พ.ศ. 2546 ด้วยปัญหาที่สำคัญ 3 ประการ คือ (1) งานวิจัยประกอบด้วยงานที่สำคัญ 2 ส่วนที่เป็นอิสระจากกัน ได้แก่การสร้างและพัฒนาแบบทดสอบ CULI-TEPOIC จำนวน 3 ชุด และการสร้างและพัฒนาโปรแกรมคอมพิวเตอร์เพื่อการทดสอบอีก 5 โปรแกรม ทั้งนี้เพราะความต้องการของหน่วยงานในขณะนั้นเป็นเช่นนี้ (2) งานแต่ละส่วนใช้เวลาดำเนินการนานมากเนื่องจากมีหลายขั้นตอน และ (3) การเขียนรายงานที่งานวิจัยมีลักษณะวิวิธพันธ์เช่นนี้ใช้เวลานานมากในการหารูปแบบที่เหมาะสม จึงทำให้รายงานเรื่องนี้เสร็จล่าช้ากว่ากำหนดเดิมมาก

แต่อย่างไรก็ตาม ผู้วิจัยใคร่ขอขอบคุณสถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย และสภาวิจัยแห่งชาติที่สนับสนุนให้ผู้วิจัยทำการวิจัยเรื่องนี้ทั้งเรื่องเวลา สถานที่ สิ่งอำนวยความสะดวกต่างๆ และด้านงบประมาณ จนทำให้งานวิจัยนี้สำเร็จได้ในที่สุด

ผู้วิจัยขอขอบคุณคณะกรรมการและผู้ที่เกี่ยวข้องในการเขียนข้อทดสอบ CULI-TEPOIC ข้าราชการทุกคนที่ช่วยในการผลิตแบบทดสอบและคุมสอบ โดยเฉพาะอาจารย์และบุคลากรทุกคนของศูนย์ทดสอบของสถาบันภาษาที่มีส่วนช่วยงานวิจัยนี้เป็นอย่างมากในการพัฒนาแบบทดสอบ และขอขอบคุณคุณคุณคมสัน เอี่ยมจรัสเป็นอย่างมากที่ให้ความร่วมมือเป็นอย่างดีในการพัฒนาโปรแกรมคอมพิวเตอร์จนสามารถทำงานได้ถูกต้องและครบถ้วนทุกโปรแกรม

ท้ายที่สุด ผู้วิจัยขอขอบคุณคณะกรรมการวิจัยของสถาบันภาษาทุกคนเป็นอย่างยิ่งที่ได้ช่วยกรุณาอ่านต้นฉบับรายงานการวิจัยนี้อย่างละเอียด และให้ข้อเสนอแนะหลายอย่างเพื่อให้ผู้วิจัยปรับปรุงแก้ไขให้รายงานนี้มีความถูกต้องและสมบูรณ์มากยิ่งขึ้น

ด้วยความขอบคุณยิ่ง

รศ. ดร. สุพัฒน์ สุกมลสันต์

วันที่ 18 สิงหาคม พ.ศ. 2549

บทคัดย่อภาษาไทย

ปัจจุบันนี้เป็นยุคของเทคโนโลยีเพื่อการสื่อสาร และคนทั่วโลกนิยมใช้ภาษาอังกฤษเป็นสื่อในการค้นคว้าและติดต่อกันในระดับนานาชาติ ดังนั้นภาษาอังกฤษจึงเป็นสิ่งที่จำเป็นสำหรับนิสิตนักศึกษาในยุคนี้และการทดสอบเป็นเครื่องมืออย่างหนึ่งในการประเมินสมรรถภาพทางภาษาของบุคคลดังกล่าว งานวิจัยเรื่องนี้จึงมีวัตถุประสงค์หลักที่สำคัญสองประการคือ เพื่อสร้างและพัฒนาแบบทดสอบวัดสมรรถภาพทางภาษาอังกฤษเพื่อวิชาชีพและการสื่อสารนานาชาติของนิสิตนักศึกษาและบุคคลทั่วไปจำนวน 3 ชุด และเพื่อสร้างและพัฒนาโปรแกรมคอมพิวเตอร์สำหรับการพัฒนาแบบทดสอบและการทดสอบจำนวน 5 โปรแกรม พลวิจัยได้แก่นิสิต นักศึกษาและบุคคลทั่วไปที่สนใจในการทดสอบสมรรถภาพทางภาษาอังกฤษของตนเองจำนวน 116, 124 และ 120 คน สำหรับแบบทดสอบ CULI-TEPOIC ชุดที่ 1, 2 และ 3 ที่สร้างขึ้นตามลำดับ แล้วผู้วิจัยนำผลการทดสอบมาวิเคราะห์ข้อทดสอบรายข้อโดยใช้รูปแบบประเพณีนิยมและแบบราช (Rasch Model) และสุ่มเลือกพลวิจัยเหล่านี้เป็นระบบจำนวนชุดละ 25 คนเพื่อสอบแบบทดสอบ TOEIC ส่วนการพัฒนาโปรแกรมคอมพิวเตอร์นั้น ผู้เขียน โปรแกรมอาศัยรหัสต้นฉบับ (source codes) ทั้งที่เป็นสำเนาอิเล็กทรอนิกส์ (soft copy) และสำเนาแสดงผล (hard copy) เพื่อพัฒนาโปรแกรมทั้ง 5 โปรแกรมจากที่ใช้งานได้ในระบบ DOS ให้เป็นระบบ Windows ผลการวิจัยสรุปได้ดังนี้

1. แบบทดสอบ CULI-TEPOIC ทั้ง 3 ชุดที่สร้างและพัฒนาขึ้นเป็นแบบทดสอบที่มีคุณภาพสูง เช่น มีความยากโดยเฉลี่ยในระดับปานกลาง ($p = 0.537-0.579$) มีความตรงรายข้อในระดับค่อนข้างสูง ($r_{pb} = 0.292-0.353$) มีความเที่ยงในระดับสูง ($KR_{20} = 0.893-0.930$) และมีความตรงร่วมสมัยกับแบบทดสอบที่มีชื่อเสียงคือแบบทดสอบ TOEIC ในระดับสูงและสูงมาก ($r_{xy} = 0.904-0.976$) เป็นต้น และ

2. โปรแกรมที่สร้างและพัฒนาขึ้นทั้ง 5 โปรแกรมสามารถทำงานได้ถูกต้อง และครบถ้วนเหมือนเช่นเดียวกับโปรแกรมเดิม และมีหนึ่งโปรแกรมที่ใช้งานได้สะดวกกว่า และดีกว่าของเดิม

Abstract

At present, we are in the Information and Communication Technology Age and English is used for searching for information and communication internationally. Therefore, it is necessary for the students in this age to know English and testing is a means to evaluate their proficiency. The main purposes of this study were then to construct and develop 3 forms of a "Chulalongkorn University Language Institute Test of English Proficiency for Occupational and International Communication Purposes (CULI-TEPOIC), and to construct and develop 5 computer programs for developing tests. The subjects in the study were 116, 124 and 120 testees taking CULI-TEPOIC forms 1, 2 and 3 respectively. They were students and any persons interested in testing their own English proficiency. Their answers and test items were then analyzed by both classical and Rasch models. Later, 25 testees from each group were randomized systematically to take a TOEIC test. As for the computer program development, source codes from soft and hard copies of the existing programs running under DOS system were used to transfer them to work under Windows system. The findings can be summarized as follows:

1. Generally speaking, the 3 forms of CULI-TEPOIC have high qualities. For example, on average their difficulty indexes are moderate ($p=0.537-0.579$), their test item validity indexes are rather high ($r_{pb} = 0.292-0.353$), their reliability coefficients are high ($KR_{20} = 0.893-0.930$) and their concurrent validity indexes with TOEIC, a well-known test, are high and very high ($r_{xy} = 0.904-0.976$).
2. All 5 newly-constructed computer programs can work as accurately and completely as the old programs, and one of them is even more user-friendly and better.

สารบัญเรื่อง

	หน้าที่
คำนำและกิตติกรรมประกาศ	ii
บทคัดย่อ ภาษาไทย	iii
บทคัดย่อภาษาอังกฤษ	iv
บทที่ 1: บทนำ	
• ความสำคัญและที่มาของปัญหาการวิจัย	1
• วัตถุประสงค์ของ โครงการวิจัย	3
• ขอบข่ายของการวิจัย	4
• ข้อตกลงเบื้องต้น	5
• ข้อจำกัดของการวิจัย	5
• คำนิยามศัพท์ที่ใช้ในการวิจัยครั้งนี้	6
• ประโยชน์ที่คาดว่าจะได้รับ และหน่วยงานที่จะนำผลการวิจัยไปใช้	9
บทที่ 2: วรรณกรรมและเอกสารที่เกี่ยวข้อง	
• ความรู้ทั่วไปเกี่ยวกับการทดสอบวัดสมิททิภาพทั่วไปทางภาษาอังกฤษ	11
• การสร้างและพัฒนาแบบทดสอบสมิททิภาพทั่วไปทางภาษา	12
• ความรู้เบื้องต้นเกี่ยวกับรูปแบบบราช	35
• ความสำคัญและลักษณะเฉพาะของ โปรแกรมที่จะพัฒนาขึ้นใช้	41
• งานวิจัยที่เกี่ยวข้องกับการสร้างและพัฒนา แบบทดสอบสมิททิภาพทั่วไปทางภาษาอังกฤษ	61
บทที่ 3: วิธีดำเนินการวิจัย	
• ส่วนที่ 1 การสร้างและพัฒนาแบบทดสอบ CULI-TEPPOIC	62
• ประชากรและพลวิจัย	62
• เครื่องมือการวิจัย (และการพัฒนา)	63
• การวิเคราะห์ข้อมูล	64
• สถิติที่ใช้	66
• ส่วนที่ 2 การสร้างและพัฒนาโปรแกรมคอมพิวเตอร์ที่ใช้ในการพัฒนาแบบทดสอบ	69
• รายละเอียดเกี่ยวกับการสร้างและพัฒนาโปรแกรมคอมพิวเตอร์ทั้ง 5 โปรแกรม	69
• การทดสอบการใช้โปรแกรมที่สร้างและพัฒนาขึ้น	76

บทที่ 4: ผลการวิจัย

- ส่วนที่ 1 การสร้างและพัฒนาแบบทดสอบ CULI-TEPPOIC 96
 - ค่าสถิติพื้นฐานของแบบทดสอบแต่ละชุด 96
 - ค่าสถิติโดยสรุปของแยกทดสอบทั้ง 3 ชุด 97
 - ค่าความตรงร่วมสมัยของแบบทดสอบทั้ง 3 ชุดและสมการพยากรณ์ 98
 - ค่าความยาก (β) ของข้อทดสอบทั้ง 3 ชุด 99
- ส่วนที่ 2 การสร้างและพัฒนาโปรแกรมคอมพิวเตอร์ที่ใช้ในการพัฒนาแบบทดสอบ 99
 - การเปรียบเทียบผลการทำงานของโปรแกรม Classical กับ CTIA 7.0 99
 - การเปรียบเทียบผลการทำงานของโปรแกรม Rasch กับ BICAL 7.0 103
 - การเปรียบเทียบผลการทำงานของโปรแกรม Item Bank กับ ITB Pro 106
 - การเปรียบเทียบผลการทำงานของโปรแกรม CCAT กับ CCAT Pro 108
 - การเปรียบเทียบผลการทำงานของโปรแกรม Grading and T-Score กับ Grading and NTScore 111

บทที่ 5: สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

- ความนำ 111
- วัตถุประสงค์ของโครงการวิจัย 112
- วิธีการวิจัย 116
- สรุปผลการวิจัย 123
- อภิปรายผล 130
- ข้อเสนอแนะ 132

บรรณานุกรม

- ภาคผนวก ก. ตัวอย่างผลการวิเคราะห์ข้อทดสอบแบบประเพณีนิยม 140
- ภาคผนวก ข. ตัวอย่างผลการวิเคราะห์ข้อทดสอบแบบบราซ (Rasch Model) 171
- ภาคผนวก ค. ผลการวิเคราะห์ข้อทดสอบรายข้อของแบบทดสอบทั้ง 3 ชุด 196
- ภาคผนวก ง. ตารางเปรียบเทียบคะแนนผลการสอบแบบทดสอบทั้ง 3 ชุดกับ TOEIC 203
- ภาคผนวก จ. สำเนาปกของแบบทดสอบ CULI-TEPPOIC ทั้ง 3 ชุด 206
- ภาคผนวก ฉ. ความสำคัญของภาษาอังกฤษในยุคของเทคโนโลยีเพื่อการสื่อสาร 213

บทที่ 1

บทนำ

ความสำคัญและที่มาของปัญหาของการวิจัย

ปัจจุบันนี้เป็นที่ยอมรับกันทั่วไปว่าเป็นยุคของเทคโนโลยีเพื่อการสื่อสารนานาชาติ (Information and Communication Technology Age) ที่มีความก้าวหน้าหลาย ๆ ด้าน เช่น วิทยาการคอมพิวเตอร์การสื่อสารนานาชาติ การทำธุรกิจผ่าน ทางระบบ Internet และการเรียนรู้ผ่านทางระบบเครือข่ายคอมพิวเตอร์(Internet) เป็นต้น นอกจากนี้ คนทั่วโลกนิยมใช้ภาษาอังกฤษเป็นสื่อสากลเพื่อการสื่อสารนานาชาติหรือสืบค้นข้อมูลต่าง ๆ ที่เกี่ยวข้องทั้งทางระบบเครือข่ายคอมพิวเตอร์ หรือสื่ออื่น ๆ มากที่สุดในโลก กล่าวคือ ประมาณร้อยละ 30 ของผู้ใช้ระบบเครือข่ายคอมพิวเตอร์หรือประมาณ 312,924,679 ล้านคนใช้ภาษาอังกฤษ (Internet World Stat, 2006) นอกจากนี้ภาษาอังกฤษยังเป็นภาษาที่ใช้ในสิ่งพิมพ์มากที่สุดในโลก รวมทั้งเป็นภาษาซึ่งผู้ที่ไม่ใช่เจ้าของภาษาใช้เพื่อพูดเป็นจำนวนที่มากที่สุดในโลกด้วย (Vistawide, 2006) ด้วยเหตุนี้จึงเป็นที่ยอมรับกันทั่วไปด้วยว่าภาษาอังกฤษมีความจำเป็นต่อการศึกษาค้นคว้าของนิสิตนักศึกษาและคนทั่วไป รวมทั้งมีความสำคัญต่อการประกอบธุรกิจของนักธุรกิจทั่วไปและการประกอบอาชีพของผู้ที่อยู่ในวงการอื่น ๆ อีกด้วย เช่น แพทย์ พยาบาล ทหาร ตำรวจ วิศวกร นักกฎหมายและบุคคลอื่น ๆ เป็นต้น

มหาวิทยาลัยมีหน้าที่และความรับผิดชอบ โดยตรงต่อสังคมในการช่วยสร้างบัณฑิตให้มีความรู้ ความพร้อม และความสามารถทันต่อการเปลี่ยนแปลงของสังคม ทั้งนี้เพื่อให้เป็นไปตามปณิธานของมหาวิทยาลัยที่จะสร้างบัณฑิตให้มีความรู้ ความสามารถ คิดเป็นและทำเป็นเพื่อออกไปปรับใช้สังคม

จากการวิจัยเชิงสำรวจจำนวนมาก มักพบว่านักศึกษา บัณฑิตและหน่วยงานต่าง ๆ ของราชการ องค์กรของรัฐวิสาหกิจ และบริษัทห้างร้านเอกชน มีความต้องการและมีการใช้ภาษาอังกฤษในการบริหารกิจการในระดับปานกลางถึงในระดับสูง โดยเฉพาะอย่างยิ่งผู้ที่เป็นนักวิชาการและบุคคลในระดับบริหาร กิจการต้องใช้อังกฤษเพื่อการอ่าน เขียน ฟัง และพูดในระดับมากตามลำดับ ปริมาณความต้องการเช่นนี้ มีมานานแล้ว และมีแนวโน้มว่าจะมีมากขึ้นเรื่อย ๆ (อัจฉรา วงศ์โสธร และคณะ, 2523:20; บังอร สว่างโรรส, 2527:38; สุพัฒน์ สุขมลสันต์ และคณะ, 2529:63-68)

ดังนั้น ในปี พ.ศ. 2533 จุฬาลงกรณ์มหาวิทยาลัย ได้เล็งเห็นความจำเป็นในการสื่อสารนานาชาติระหว่างประเทศโดยใช้ภาษาอังกฤษเป็นสื่อกลางเป็นอย่างมาก โดยเชื่อว่า หากบัณฑิตของจุฬาลงกรณ์มหาวิทยาลัยมีความสามารถทั้งทางเนื้อหาวิชาการ และการใช้ภาษาอังกฤษในระดับที่สามารถติดต่อทำความเข้าใจกับชาวต่างประเทศได้ ก็จะทำให้บัณฑิตมีคุณค่าเป็นที่ต้องการของตลาดแรงงานและสังคมมากขึ้น ซึ่งจะทำให้เป็นแรงผลักดันให้สังคมไปสู่สังคมอุตสาหกรรมได้ทันเวลาและปลอดภัย ดังนั้น ฝ่ายวิชาการของ

มหาวิทยาลัยจึงเห็นความจำเป็นที่จะต้องปรับความสามารถในการใช้ภาษาอังกฤษของบัณฑิตให้ได้มาตรฐานที่ทางมหาวิทยาลัยกำหนด และวิธีหนึ่งที่จะทำได้โดยการทดสอบสมิทริภาพทั่วไปทางภาษาอังกฤษของนิสิตชั้นปีที่ 3-4 หรือ 2 ปีสุดท้ายของหลักสูตรต่าง ๆ เพื่อเป็นการกระตุ้นให้เกิดการตื่นตัวทั้งอาจารย์และนิสิตให้สนใจและให้ความสำคัญต่อการเรียนการสอนภาษาอังกฤษด้วยวิธีการต่าง ๆ (อัจฉรา และคณะ, 2535; สุพัฒน์ และคณะ, 2533)

ในปี พ.ศ. 2532 ฝ่ายวิชาการของมหาวิทยาลัย จึงได้ร่วมกับฝ่ายวิจัยและพัฒนาของสถาบันภาษา ทำการวิจัยและพัฒนาแบบทดสอบวัดสมิทริภาพทั่วไปทางภาษาอังกฤษขึ้น เป็นโครงการ 3 ปี คือ ปี พ.ศ. 2533-2536 ทั้งนี้เพราะมหาวิทยาลัยยังขาดแบบทดสอบมาตรฐานที่จะใช้ให้ได้ประโยชน์ตามที่ต้องการ แบบทดสอบที่สร้างขึ้นตามโครงการดังกล่าวมีชื่อว่า แบบทดสอบสมิทริภาพทางภาษาอังกฤษของจุฬาลงกรณ์มหาวิทยาลัย (Chulalongkorn University Test of English Proficiency) หรือที่รู้จักกันทั่วไปในปัจจุบันนี้คือ CU-TEP แบบทดสอบนี้สร้างขึ้นตามลักษณะของแบบทดสอบวัดสมิทริภาพทางภาษาอังกฤษที่เป็นที่รู้จักกันทั่วไป ชื่อ TOEFL (Test of English as a Foreign Language) ของ ETS (Educational Testing Service) ของอเมริกา กล่าวคือ แบบทดสอบ CU-TEP มีข้อทดสอบแบบปรนัย 120 ข้อ ประกอบด้วย ข้อทดสอบการฟังเข้าใจความ 35 ข้อ ข้อทดสอบการอ่านเข้าใจความ 60 ข้อ ข้อทดสอบการเขียน 25 ข้อ และเป็นข้อทดสอบการเขียนแบบเรียงความอีก 3 ข้อ รวมเป็นข้อทดสอบทั้งหมด 123 ข้อ และใช้เวลาสอบ 3 ชั่วโมง (สุพัฒน์ สุกมลสันต์และคณะ, 2533)

อนึ่ง ตั้งแต่ปี พ.ศ. 2533 เป็นต้นมา สถาบันภาษาได้สร้างและพัฒนาแบบทดสอบ CU-TEP ปีละ 1-4 ชุดเพื่อให้บริการการทดสอบแก่นิสิต นักศึกษา และบุคคลทั่วไปที่สนใจ โดยเฉพาะอย่างยิ่งสำหรับผู้ประสงค์จะสอบเข้าศึกษาในระดับบัณฑิตศึกษาที่จุฬาลงกรณ์มหาวิทยาลัย และบางองค์กรที่ต้องการคัดเลือกบุคลากรใหม่หรือคัดเลือกบุคลากรไปศึกษาต่อ แบบทดสอบที่สร้างขึ้นแต่ละชุดเป็นแบบทดสอบ ที่มีคุณภาพที่ดีเนื่องจากมีกระบวนการในการตรวจสอบคุณภาพและความถูกต้องตามหลักการวิจัยและการพัฒนาแบบทดสอบที่ดี เช่น แบบทดสอบ CU-TEP ในปี พ.ศ. 2533 ทำการทดสอบกับนิสิตระดับปริญญาโท-เอก และบุคคลภายนอกที่สนใจ รวมทั้งหมด 1,684 คน ปรากฏว่า แบบทดสอบมีความเที่ยงสูงมาก คือ $KR_{20} = 0.90$ มีความยากง่ายปานกลาง ($\Delta=13.45$) มีค่าอำนาจจำแนกโดยเฉลี่ยอยู่ในระดับที่ดี ($r_{pb} = 0.36$) และผู้ชำนาญการทางทดสอบ พิจารณาแล้วว่ามีตรงเชิงเนื้อหา (สุพัฒน์ สุกมลสันต์และคณะ, 2533:4,26) และแบบทดสอบ CU-TEP จำนวน 4 ชุดในปี พ.ศ. 2534 เป็นแบบทดสอบที่มีคุณภาพดีมาก 2 ชุด และระดับดีอีก 2 ชุด กล่าวคือ มีความเที่ยงอยู่ในระดับสูง ($KR_{20} = 0.854-0.910$ และ $\phi = 0.822 - 0.896$) มีความยากง่ายปานกลาง ($\Delta=12.831-13.142$) และอำนาจจำแนกอยู่ในระดับใช้ได้ ($r_{pb} = 0.230-0.288$) (อัจฉรา วงศ์โสธร และคณะ, 2535:27) และแบบทดสอบ CU-TEP ที่สร้างในปี พ.ศ. 2535 จำนวน 3 ชุด ที่ได้ทำการทดสอบกับ

นิสิต นักศึกษา และบุคคลภายนอกที่สนใจจำนวน 447, 463 และ 657 คนตามลำดับ ปรากฏว่า แบบทดสอบ ทั้ง 3 ชุด มีความเที่ยงสูง ($KR_{20} = 0.882-0.915$) มีความยากง่าย ($\Delta = 13.183-13.639$) และมีอำนาจจำแนกในระดับที่ดี ($r_{bis} = 0.325-0.378$) (สุพัฒน์ สุขมลสันต์, 2536:1-3) เป็นต้น

อนึ่ง ตั้งแต่ พ.ศ. 2535 เป็นต้นมา สถาบันภาษาได้ให้บริการทดสอบสมรรถภาพทางภาษาอังกฤษโดยใช้แบบทดสอบ CU-TEP แก่บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย เพื่อคัดเลือกนิสิตเข้าศึกษาในระดับบัณฑิตศึกษาและให้บริการทดสอบภาษาอังกฤษแก่หน่วยงานอื่น ๆ ทั้งภายนอกและภายในมหาวิทยาลัย รวมทั้งหน่วยงานในภาครัฐ ภาคเอกชน รัฐวิสาหกิจ และบุคคลทั่วไปประมาณปีละ 70 ครั้ง และมีผู้มารับบริการประมาณปีละ 13,000 คน (รายงานประจำปี พ.ศ. 2542) ปัจจุบันนี้ผลการทดสอบของแบบทดสอบ CU-TEP เป็นที่ยอมรับกันทั่วไปในมหาวิทยาลัย และนอกมหาวิทยาลัยในระดับหนึ่งแล้ว จนทำให้บางมหาวิทยาลัยมีแบบทดสอบในทำนองเดียวกันเกิดขึ้น แต่ยังไม่แพร่หลายนัก ในปัจจุบันนี้ แม้ว่าสถาบันภาษาจะยังเป็นผู้สร้างและพัฒนาแบบทดสอบ CU-TEP แต่การจัดสอบและบริการการสอนเป็นหน้าที่ของศูนย์ทดสอบแห่งจุฬาลงกรณ์มหาวิทยาลัย (Testing Center of Chulalongkorn University)

แต่อย่างไรก็ตาม เนื่องจากแบบทดสอบ CU-TEP วัดสมรรถภาพทั่วไปทางภาษาอังกฤษสำหรับผู้ที่ต้องการวัดความสามารถทั่ว ๆ ไปทางภาษาอังกฤษของตนเอง จึงยังไม่เหมาะสำหรับการทดสอบความสามารถเฉพาะทางเพื่อวิชาชีพและการสื่อสารนานาชาติเป็นชาติ ซึ่งเป็นแบบทดสอบที่สังคมมีความต้องการมากเพิ่มขึ้น ดังจะเห็นได้จากมีแบบทดสอบที่นิยมใช้ทั่วไปขณะนี้ คือแบบทดสอบ TOEIC (Test of English for International Communication) ของ ETS (Educational Testing Service) แห่งอเมริกา แต่ว่าค่าบริการทดสอบของแบบทดสอบนี้ยังสูงมากสำหรับคนทั่วไป

ดังนั้น สถาบันภาษาจึงเห็นความสำคัญที่จะจัดให้มีโครงการวิจัยเรื่อง การพัฒนาแบบทดสอบวัดสมรรถภาพทางภาษาอังกฤษเพื่อวิชาชีพและการสื่อสารนานาชาติของสถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย หรือ CULI-TEPOIC [Chulalongkorn University Language Institute Test of English Proficiency for Occupational and International Communication Purposes] ให้ได้มาตรฐานขึ้น และทำการพัฒนาโปรแกรมคอมพิวเตอร์ที่จำเป็นต้องใช้เพื่อการพัฒนาข้อทดสอบและการทดสอบให้ทันสมัยยิ่งขึ้น ทั้งนี้ เพื่อให้ระบบการทดสอบของสถาบันภาษามีความสมบูรณ์ยิ่งขึ้นด้วย

วัตถุประสงค์ของโครงการวิจัย

1. เพื่อสร้างและพัฒนาแบบทดสอบวัดสมรรถภาพทางภาษาอังกฤษเพื่อวิชาชีพและการสื่อสารนานาชาติของสถาบันภาษาจุฬาลงกรณ์มหาวิทยาลัย ให้มีคุณภาพและได้มาตรฐาน จำนวน 3 ชุด

2. เพื่อพัฒนาโปรแกรมคอมพิวเตอร์ที่ใช้ในการพัฒนาแบบทดสอบและการทดสอบให้ทันสมัย 5 โปรแกรม

ขอบข่ายของการวิจัย

1. แบบทดสอบที่สร้างและพัฒนาเป็นแบบทดสอบวัดสมรรถภาพทางภาษาอังกฤษเพื่อวิชาชีพ และการสื่อสารนานาชาติจำนวน 3 ชุดๆละ 100 ข้อในลักษณะที่คล้ายคลึงกับแบบทดสอบ TOEIC กล่าวคือแบบทดสอบแต่ละชุดประกอบด้วยดังนี้

ก. การฟัง (Listening) จำนวน 50 ข้อ ประกอบด้วย

Part 1: ฟังข้อความประกอบรูปภาพ (Photographs) จำนวน 10 ข้อ

Part 2: ฟังคำถามและตอบ (Question – Response) จำนวน 10 ข้อ

Part 3: ฟังข่าวและประกาศ (News and Announcements) จำนวน 10 ข้อ

Part 4: ฟังบทสนทนาสั้นๆ (Short Conversations) จำนวน 10 ข้อ

Part 5: ฟังการพูดสั้นๆ (Short Talks) จำนวน 10 ข้อ

ข. การเขียนและอ่าน (Writing and Reading) จำนวน 50 ข้อ ประกอบด้วย

Part 6: อ่านและเติมคำในประโยค (Incomplete Sentences) จำนวน 15 ข้อ

Part 7: อ่านหาข้อผิดพลาด (Error Recognition) จำนวน 15 ข้อ

Part 8: อ่านเข้าใจความ (Reading Comprehension) จำนวน 20 ข้อ

ค. ระยะเวลาในการทดสอบทั้งหมดรวม 1.40 ชั่วโมง คือ

การฟัง (ก) ใช้เวลา 40 นาที

การอ่าน (ข) ใช้เวลา 60 นาที

2. แบบทดสอบที่สร้างและพัฒนาเป็นแบบทดสอบสำหรับนิสิต/นักศึกษาในระดับมหาวิทยาลัยและบุคคลทั่วไปที่สนใจที่เน้นเรื่องการสื่อสารนานาชาติเพื่อวิชาชีพและการสื่อสารนานาชาติ
3. โปรแกรมคอมพิวเตอร์ที่สร้างขึ้นเพื่อใช้ในการพัฒนาแบบทดสอบและการทดสอบมีจำนวน 5 โปรแกรม คือ โปรแกรมวิเคราะห์ข้อทดสอบแบบประเพณีนิยม (Classical Model)

- โปรแกรมวิเคราะห์ข้อทดสอบแบบราช (Rasch Model) โปรแกรมธนาคารข้อทดสอบ (Item Banking) โปรแกรมการทดสอบแบบปรับเปลี่ยน (Adaptive Testing) และโปรแกรมกำหนดระดับคะแนน (Grade Assignment)
4. โปรแกรมคอมพิวเตอร์ที่สร้างขึ้นทั้งหมดเป็นการพัฒนามาจากโปรแกรมเดิมที่ใช้ในระบบ DOS (Disk Operation System) เป็นระบบ Windows (Window Operation System)

ข้อตกลงเบื้องต้น

ในการวิจัยครั้งนี้ ผู้วิจัยมีข้อตกลงเบื้องต้น (หรือข้อสันนิษฐานเบื้องต้น) ดังต่อไปนี้ คือ

1. ผู้สอบแบบทดสอบที่ใช้ในการวิจัยครั้งนี้ได้พยายามมากที่สุดในการตอบแบบทดสอบเพื่อให้ได้คะแนนมากที่สุดเท่าที่จะสามารถทำได้ เพราะการทดสอบแต่ละครั้งเป็นการทดสอบจริงเพื่อทดสอบความรู้และความสามารถของตนเองว่ามีมากน้อยเพียงใด หรือเพื่อนำผลการทดสอบไปประกอบการพิจารณาสมัครเข้าทำงานของตนเอง ซึ่งเป็นความต้องการของผู้สอบแต่ละคน
2. โปรแกรมคอมพิวเตอร์ทั้ง 5 โปรแกรมที่ใช้ในระบบ DOS เป็นโปรแกรมที่ทำงานได้ผลการคำนวณถูกต้องและเชื่อถือได้ เนื่องจากเป็นโปรแกรมที่ได้พัฒนาและตรวจสอบผลการทำงานมานานแล้วว่าถูกต้องและเชื่อถือได้ (แต่ที่ต้องการปรับปรุงและพัฒนาให้สามารถทำงานได้ในระบบ Windows เนื่องจากปัจจุบันนี้เครื่องคอมพิวเตอร์โดยทั่วไปไม่ได้ใช้ระบบ DOS แล้ว)

ข้อจำกัดของการวิจัย

การวิจัยนี้มีข้อจำกัดบางอย่างดังนี้

1. แบบทดสอบที่ใช้ในการวิจัยครั้งนี้มี 3 ชุด คือ CULI-TEPOIC ชุด 1 ชุด 2 และชุด 3 แบบทดสอบทั้ง 3 ชุดนี้ สร้างจากตารางกำหนดลักษณะเฉพาะของข้อทดสอบ (Table of Test Specification) เดียวกัน แต่ทำการทดสอบกับกลุ่มผู้สอบที่แตกต่างกัน แม้ว่าจะอยู่ในระดับอุดมศึกษาและบุคคลทั่วไปเช่นเดียวกันก็ตาม ดังนั้น ค่าสถิติที่ได้จากการวิเคราะห์ข้อทดสอบแต่ละชุดอาจแตกต่างกันบ้าง
2. การคำนวณหาความตรงร่วมสมัย (concurrent validity) ของแบบทดสอบที่สร้างขึ้นทั้ง 3 ชุดอาศัยคะแนนผลการสอบจากแบบทดสอบ TOEIC กับคะแนนผลการสอบจากแบบทดสอบที่สร้างขึ้น ผู้วิจัยได้สุ่มเลือกผู้ที่สอบได้คะแนนในระดับต่างๆอย่างมีระบบจำนวนชุดละ 25 คนเพื่อให้ไปสอบแบบทดสอบ TOEIC แต่ปรากฏว่ามีผู้ไปสอบเพียงชุดละ 19-24 คนเท่านั้น

ซึ่งเป็นจำนวนที่ค่อนข้างน้อย ดังนั้น สมการพยากรณ์คะแนนผลการทดสอบแบบทดสอบ TOEIC จึงยังมีความคลาดเคลื่อนค่อนข้างมาก

คำนิยามศัพท์ที่ใช้ในการวิจัยครั้งนี้

1. **เกณฑ์ความสามารถ** หมายถึงแถบความสามารถ (ability band) เชิงพฤติกรรมที่บ่งบอกว่าผู้สอบที่มีความรู้ความสามารถในแต่ละระดับสามารถทำอะไรได้บ้าง เช่น ผู้ที่มีความสามารถสูงด้านการอ่านสามารถอ่านเข้าใจความในด้านใดบ้างและมากน้อยเพียงใด เป็นต้น
2. **นิสิต/นักศึกษาไทย** หมายถึงผู้สอบซึ่งเป็นคนไทยที่มีความรู้ความสามารถในระดับอุดมศึกษาและสมัครสอบแบบทดสอบวัดสมรรถภาพทางภาษาอังกฤษเพื่อวิชาชีพและการสื่อสารนานาชาตินานาชาติของสถาบันภาษาจุฬาลงกรณ์มหาวิทยาลัย หรือ CULI-TEPOIC [Chulalongkorn University Language Institute Test of English Proficiency for Occupational and International Communication Purposes] เพื่อการศึกษาต่อในประเทศหรือต่างประเทศ บุคคลเหล่านี้มีระดับความสามารถทางภาษาอังกฤษแตกต่างกัน และมีภูมิหลังทางการศึกษา สังคม และเศรษฐกิจต่างกัน
3. **แบบทดสอบสมรรถภาพ (Proficiency Test)** หมายถึงแบบทดสอบที่ประกอบด้วยข้อทดสอบที่มุ่งทดสอบความรู้ความสามารถทั่วไปที่จำเป็นต่อการกระทำการอย่างใดอย่างหนึ่งในอนาคต เช่น เพื่อการศึกษาต่อ เพื่อการสื่อสารนานาชาติ และเพื่อใช้ในวิชาชีพ เป็นต้น และเนื้อหาเหล่านี้ไม่ได้กำหนดไว้ในหลักสูตรของรายวิชาใดวิธีหนึ่งโดยเฉพาะ
4. **สมรรถภาพทั่วไปทางภาษาอังกฤษเพื่อวิชาชีพและการสื่อสารนานาชาติ** หมายถึงความรู้ความสามารถของผู้สอบที่ได้จากคะแนนสอบของแบบทดสอบวัดสมรรถภาพทางภาษาอังกฤษเพื่อวิชาชีพและการสื่อสารนานาชาติของสถาบันภาษาจุฬาลงกรณ์มหาวิทยาลัย หรือ CULI-TEPOIC [Chulalongkorn University Language Institute Test of English Proficiency for Occupational and International Communication Purposes] คะแนนเหล่านี้บ่งบอกถึงความรู้ความสามารถทั่วไปทางภาษาอังกฤษของผู้สอบที่ไม่ได้ระบุไว้ในรายวิชาใดวิชาหนึ่งโดยเฉพาะ แต่มีความจำเป็นต่อการสื่อสารนานาชาติในวิชาชีพและการสื่อสารนานาชาติทั่วไป

5. แบบทดสอบผลสัมฤทธิ์ (Achievement Test) หมายถึงแบบทดสอบที่สามารถให้ข้อมูลที่บ่งบอกได้ว่าผู้เรียนรายวิชาใดรายวิชาหนึ่งได้เรียนรู้เนื้อหาของรายวิชาดังกล่าวมากน้อยเพียงใด และมีลักษณะเฉพาะดังนี้
 1. หน้าที่ เพื่อใช้ทดสอบว่าผู้เรียนได้เรียนรู้อะไรบ้าง
 2. เนื้อหา มีเนื้อหาครอบคลุมหลักสูตรใดหลักสูตรหนึ่ง
 3. ลักษณะข้อทดสอบ เน้นหนักที่ข้อทดสอบวิวิธพันธุ์ (Heterogeneous Item) คือ ประกอบด้วยข้อทดสอบที่มุ่งทดสอบสิ่งต่าง ๆ แยกต่างหากและมีจำนวนมาก
7. แบบทดสอบผลสัมฤทธิ์ทั่วไป (General Achievement Test) หมายถึงแบบทดสอบผลสัมฤทธิ์ชนิดหนึ่งซึ่งมีลักษณะเฉพาะดังนี้
 1. หน้าที่ เพื่อใช้วัดผลสัมฤทธิ์ในการเรียนระดับ “มหภาษา” (macro-language)
 2. โครงสร้าง ประกอบด้วย “มหภาษา”
 3. การทดสอบ ทดสอบด้วยข้อทดสอบแบบวัดจุดย่อย (discrete-point item) และข้อทดสอบแบบวัดรวมสรุป (global item) ซึ่งมีได้มุ่งหวังว่าจะทดสอบความรู้แต่ละจุดย่อยของเนื้อหา เช่น การเขียนเรียงความ และแบบทดสอบ โคลซ (cloze tests) เป็นต้น
8. ความสามารถศึกย์ (competence) หมายถึงความสามารถทั่ว ๆ ไปทางภาษาอังกฤษของผู้เรียนเกี่ยวกับความรู้และความสามารถที่จะใช้ความรู้ นั้น ซึ่งความรู้ นี้ประกอบด้วย 4 อย่าง คือ ไวยากรณ์ (grammar) จิตภาษาศาสตร์ (psycholinguistics) ความรู้ทางสังคมและวัฒนธรรม (socio-cultural knowledge) และความสามารถเพื่อใช้ภาษา (ability for use)
9. ความสามารถฉฉฉฉ (performance) หมายถึงความสามารถในการใช้ภาษาได้จริงในเหตุการณ์จริง (actual use in actual events)
10. ความสามารถศึกย์เชิงภาษาศาสตร์ (linguistic or grammatical competence) หมายถึง ความรู้ ความสามารถและทักษะในการที่จะเข้าใจความหมายของคำหรือข้อความตามตัวอักษร ซึ่งถือได้ว่าเป็นความสามารถในระดับ “จุลภาษา” (micro-language)
11. ความสามารถศึกย์เชิงสังคมภาษาศาสตร์ (sociolinguistic competence) หมายถึง ความสามารถที่จะเข้าใจและใช้ภาษาได้อย่างเหมาะสมกับสังคมแวดล้อม ทั้งด้านความหมาย และรูปแบบของภาษาตามหน้าที่ของการสื่อสารนานาชาติ (communicative functions) เช่น การออกคำสั่ง อธิบาย ขอร้องและขอโทษ เป็นต้น
12. ความสามารถศึกย์เชิงความสัมพันธ์ระหว่างรูปแบบของภาษาและความหมาย (discourse competence) หมายถึงความสามารถที่จะเข้าใจว่ารูปแบบของภาษา (grammatical form)

สัมพันธ์กับความหมายอย่างไร เช่น ข้อความต่าง ๆ สัมพันธ์กันอย่างไร กล่าวคือรู้จักการใช้การเชื่อมความ(cohesion) และความหมายของข้อความว่าสัมพันธ์กันอย่างไร หรือรู้จักการใช้การเชื่อมคำ (coherence) เป็นต้น

13. ความสามารถศักยภาพเชิงกลวิธี (strategic competence) หมายถึงความสามารถในการที่จะสื่อสารความหมายได้ทั้งการใช้ภาษา (verbal) และไม่ใช่ภาษา (non-verbal) เช่น การเน้นข้อความบางแห่งที่ต้องการ การใช้ท่าทางหรือกิริยาประกอบการสื่อสารนานาชาติ การใช้ข้อความอื่นเพื่อสื่อความหมายเดียวกันที่ตนมีปัญหา หรือการใช้รูปประโยคใหม่ (paraphrase) เป็นต้น
14. จุลภาษา (micro-language) หมายถึงความรู้พื้นฐานเกี่ยวกับระบบหน่วยเสียง หน่วยคำ วากยสัมพันธ์ และความสัมพันธ์ระหว่างระบบเหล่านี้กับระบบความหมายของคำ
15. มหาภาษา (macro-language) หมายถึงความรู้ความสามารถเกี่ยวกับในการใช้ภาษาเพื่อประโยชน์ในการสื่อสารนานาชาติ รวมทั้งการส่งและการรับสาร โดยการพูดและการเขียน หรืออาจเรียกว่าความสามารถจลน์รวมทางภาษา (integrated language performance)
16. แบบทดสอบวัดจุดย่อย (discrete-point test) หมายถึงแบบทดสอบที่มุ่งทดสอบความรู้ความสามารถระดับจุลภาษาที่ละจุดหรือที่ละปัญหา เพราะมีความเชื่อในการทดสอบว่า ข้อทดสอบที่ดีควรทดสอบปัญหาที่ละจุด (discrete points) เช่น แบบทดสอบแบบเลือกตอบ (multiple-choice test) เป็นต้น
17. แบบทดสอบวัดรวมสรุป (global test/holistic test/integrated test) หมายถึงแบบทดสอบที่มุ่งทดสอบความรู้ความสามารถหลายด้านพร้อมกัน เพราะมีความเชื่อว่าข้อทดสอบที่ดีควรทดสอบภาษาหลาย ๆ ด้าน รวมกันในเวลาเดียวกัน (holistic) เช่น แบบทดสอบ cloze test และแบบทดสอบเรียงความ เป็นต้น
18. ความสามารถระดับความรู้ (knowledge) หมายถึงความสามารถที่แสดงว่าผู้สอบสามารถจำ (recognize) และระลึก (recall) ข้อเท็จจริงต่าง ๆ ที่ได้ศึกษาแล้วได้ เช่น การสะกดคำ ความหมายของคำศัพท์ กฎเกณฑ์ทางไวยากรณ์ ความหมายของประโยค การตอบสนองต่อประโยคคำถามที่คุ้นเคยมาแล้ว
19. ความสามารถระดับถ่ายโอน (transfer) หมายถึงความสามารถที่แสดงว่าผู้สอบสามารถรับรู้และเข้าใจ (reception) ข้อเท็จจริงต่าง ๆ ที่ได้ศึกษามาแล้ว แต่ปรากฏในสถานการณ์หรือบริบทใหม่ที่แตกต่างไปจากเดิม เช่น เข้าใจข้อความที่บันทึกลงในแถบบันทึกเสียงได้ และเข้าใจเนื้อเรื่องที่อ่านซึ่งประกอบด้วยประโยคและคำศัพท์ที่เคยศึกษามาแล้วแต่จัดลำดับ

เนื้อหาใหม่ได้ เป็นต้น นอกจากนี้ผู้สอบยังสามารถนำความรู้เดิมที่ศึกษามาแล้วประยุกต์ใช้ (application) ในสถานการณ์ใหม่ได้อย่างถูกต้องด้วย เช่น นำความรู้ทางไวยากรณ์มาใช้ในการพูดหรือเขียนได้ และเขียนข้อความใหม่โดยการเปลี่ยนแปลงรูปประโยคให้เหมาะสมกับกาล (tense) ที่ต้องการสื่อสารนานาชาติได้ เป็นต้น

20. ความสามารถระดับสื่อสาร (communication) หมายถึงความสามารถที่แสดงว่าผู้สอบเมื่อฟังหรืออ่านข้อความแล้ว และสามารถเข้าใจความ (comprehension) ได้มากน้อยเพียงใด เมื่อฟังหรืออ่านข้อความแล้ว และสามารถแสดงความคิดเห็นของตน (self-expression) โดยการพูดและการเขียนได้มากน้อยเพียงใด ในระดับนี้ความคล่องแคล่ว (fluency) ในการใช้ภาษามีความสำคัญมากกว่าความถูกต้อง (accuracy) ทางไวยากรณ์
21. ความสามารถระดับวิพากษ์วิจารณ์ (criticism) หมายถึงความสามารถที่แสดงว่าผู้สอบสามารถวิเคราะห์ (analysis) ความแตกต่างของภาษาที่ใช้ ผู้ใช้ภาษา และจุดมุ่งหมายของการใช้ภาษาได้มากน้อยเพียงใด เช่น ความแตกต่างระหว่างภาษาพูดและภาษาเขียน ภาษาทางราชการและภาษาใช้ทั่วไป อารมณ์ของผู้พูดและความหมายแฝงโดยนัยของข้อความที่ใช้ เป็นต้น และต้องสามารถทำการสังเคราะห์ (synthesis) ภาษาในเชิงของการวิพากษ์วิจารณ์ได้ สามารถใช้ภาษาได้ตามจุดมุ่งหมายที่ตนมุ่งประสงค์ได้ รวมทั้งสามารถที่จะทำการประเมินผล (evaluation) ภาษาที่ผู้อื่นใช้ได้ด้วยว่าเป็นอย่างไร เหมาะสมหรือไม่ เป็นต้น
22. โปรแกรมคอมพิวเตอร์ที่ใช้ในการพัฒนาแบบทดสอบและการทดสอบ หมายถึง โปรแกรมวิเคราะห์ข้อทดสอบแบบประเพณีนิยม (Classical Model) โปรแกรมวิเคราะห์ข้อทดสอบแบบราช (Rasch Model) โปรแกรมธนาคารข้อทดสอบ (Item Banking) โปรแกรมการทดสอบแบบปรับเปลี่ยน (Adaptive Testing) และโปรแกรมกำหนดระดับคะแนน (Grade Assignment)

ประโยชน์ที่คาดว่าจะได้รับ และหน่วยงานที่จะนำผลการวิจัยไปใช้ประโยชน์

1. นิสิต/นักศึกษาและบุคคลทั่วไปเมื่อทำการทดสอบแล้วจะได้ทราบว่าระดับสมรรถภาพทั่วไปเพื่อวิชาชีพและการสื่อสารนานาชาตินานาชาติทางภาษาอังกฤษของตนมีมากน้อยเพียงใด ซึ่งจะเป็นตัวกระตุ้นให้นิสิต/นักศึกษาและบุคคลทั่วไปมีความสนใจ หรือแรงจูงใจในการเรียนภาษาอังกฤษมากยิ่งขึ้น และเป็นผลดีต่อสภาพการเรียนการสอนภาษาอังกฤษในประเทศไทยได้ทางหนึ่ง

2. สถาบันต่าง ๆ หน่วยงานของรัฐ และบริษัทห้างร้านต่าง ๆ ของเอกชนอาจใช้คะแนนผลการทดสอบประกอบการพิจารณาเพื่อคัดเลือกนิสิต/นักศึกษาและบุคคลทั่วไปรับเข้าทำงานในหน่วยงานของตน โดยไม่ต้องเสียค่าใช้จ่ายในการพัฒนาแบบทดสอบของตนเองให้สิ้นเปลืองงบประมาณ
3. คณะต่างๆของมหาวิทยาลัยอาจพิจารณาเกณฑ์ความสามารถจากระดับคะแนนสอบเป็นเกณฑ์ความสามารถขั้นต่ำของนิสิต/นักศึกษาของตนก่อนที่จะจบการศึกษา เพื่อเป็นการควบคุมคุณภาพทางภาษาของบัณฑิตของตนได้
4. ประชาชาติจะสามารถประหยัดเงินตราของประเทศได้ปีละจำนวนมาก หากว่าเกณฑ์คะแนนสอบได้รับการยอมรับทั่วไป เพราะนิสิต/นักศึกษาและบุคคลทั่วไปไม่จำเป็นต้องไปทดสอบแบบทดสอบอื่นของต่างประเทศที่ต้องเสียค่าสมัครสอบแพงมาก และในขณะเดียวกันก็เป็นการส่งเสริมการใช้ภูมิปัญญาไทยในการประหยัดเงินตราของประเทศได้ด้วย
5. สถาบันภาษาจะได้แบบทดสอบที่มีคุณภาพและได้มาตรฐานจำนวน 3 ชุด หรือ 300 ข้อ เพื่อเก็บไว้ใช้ในธนาคารข้อทดสอบสำหรับใช้ทดสอบวัดสมิทธิภาพทางภาษาอังกฤษเพื่อวิชาชีพและการสื่อสารนานาชาตินานาชาติในโอกาสต่าง ๆ ที่ต้องการทันที
6. สถาบันภาษาสามารถให้บริการทดสอบวัดสมิทธิภาพทางภาษาอังกฤษให้แก่หน่วยงานและองค์กรต่าง ๆ ที่ขอรับบริการได้ทันที ทำให้งานบริการทางวิชาการด้านทดสอบมีประสิทธิภาพและสนองความต้องการของผู้รับบริการได้มากยิ่งขึ้นนอกเหนือจากการทดสอบวัดสมิทธิภาพทั่วไปทางภาษาอังกฤษอื่นที่มีอยู่แล้ว
7. สถาบันภาษาจะได้โปรแกรมคอมพิวเตอร์จำนวน 5 โปรแกรม ซึ่งทันสมัยสำหรับใช้ในระบบการทดสอบ โดยไม่ต้องเสียเงินไปซื้อจากต่างประเทศด้วยราคาที่แพงมาก และอาจเป็นประโยชน์กับหน่วยงาน หรือองค์กรอื่นที่ดำเนินการด้านการทดสอบหรือการให้บริการทดสอบได้ด้วย

บทที่ 2

วรรณกรรมและเอกสารที่เกี่ยวข้อง

ในบทนี้ผู้เขียนจะขอกล่าวถึงวรรณกรรมและเอกสารที่เกี่ยวข้องใน 5 หัวข้อที่สำคัญ คือ

1. ความรู้ทั่วไปเกี่ยวกับการทดสอบวัดสมิทริภาพทั่วไปทางภาษาอังกฤษ
 2. การสร้างและพัฒนาแบบทดสอบสมิทริภาพทั่วไปทางภาษา
 3. ความรู้เบื้องต้นเกี่ยวกับรูปแบบราช (Rasch Model)
 4. ความสำคัญและลักษณะเฉพาะของ โปรแกรมที่จะพัฒนาขึ้นใช้
 5. งานวิจัยที่เกี่ยวข้องกับการสร้างและพัฒนาแบบทดสอบสมิทริภาพทั่วไปทางภาษาอังกฤษ
- รายละเอียดของแต่ละหัวข้อดังกล่าวมีดังนี้

ส่วนที่ 1. ความรู้ทั่วไปเกี่ยวกับการทดสอบวัดสมิทริภาพทั่วไปทางภาษาอังกฤษ

เพื่อที่จะสามารถเข้าใจการทดสอบวัดสมิทริภาพทั่วไปทางภาษาอังกฤษได้เป็นอย่างดี ผู้วิจัยขอ
นำเสนอความรู้เบื้องต้นที่เกี่ยวข้องดังต่อไปนี้ คือ

1. ทฤษฎีต่าง ๆ ทางภาษาศาสตร์และการทดสอบทางภาษา
 2. ความแตกต่างระหว่างแบบทดสอบจุดย่อยและแบบทดสอบรวมสรุป
 3. ความแตกต่างของแบบทดสอบสมิทริภาพทั่วไปกับแบบทดสอบชนิดอื่น
 4. ประเด็นปัญหาที่สำคัญของแบบทดสอบสมิทริภาพทั่วไป
 5. การสร้างและพัฒนาแบบทดสอบสมิทริภาพทั่วไปทางภาษา
- ต่อไปนี้เป็นรายละเอียดของหัวข้อต่าง ๆ ดังกล่าวแล้ว

1. ทฤษฎีต่าง ๆ ทางภาษาศาสตร์และการทดสอบทางภาษา

ทฤษฎีทางภาษามีอยู่มากมายก็จริงแต่พอที่จะแบ่งออกเป็น 3 กลุ่มใหญ่ได้ดังนี้ (Morrow, 1977:1):

- 1) กลุ่มที่นิยมการแยกภาษาออกจากสังคมอย่างเด็ดขาด (Strict-Separatists) นักภาษาศาสตร์กลุ่มนี้มี
ความเชื่อว่า ภาษามีลักษณะเหมือนกับวัตถุชิ้นหนึ่งที่แยกตัวออกจากสังคมเป็นอิสระได้อย่าง
เด็ดขาด ดังนั้นการเรียนภาษาคือการเรียนความจริงต่าง ๆ เกี่ยวกับภาษา เช่น การเรียนวรรณคดีสมัย
เก่า หรือการศึกษาที่มาของคำศัพท์ต่าง ๆ เป็นต้น โดยคิดว่าส่วนย่อย ๆ เหล่านี้ประกอบกันเป็น
ภาษา โดยไม่คำนึงถึงความแตกต่างในการใช้ภาษาในสังคมที่แตกต่างกัน นักภาษาศาสตร์กลุ่มนี้มุ่ง
“ศึกษาภาษา” (language study) นักภาษาศาสตร์กลุ่มนี้นิยมทดสอบความรู้ความสามารถของผู้เรียน
โดยการแยกภาษาออกเป็นส่วย่อยๆ เช่น การฟัง การพูด การอ่าน การเขียน และการแปล รวมทั้ง

การตีความข้อเขียนทางวรรณคดีต่างๆ เป็นต้น และนิยมใช้แบบทดสอบอัตนัย (subjective test) เป็นเครื่องมือในการทดสอบ

- 2) กลุ่มที่นิยมการแยกภาษาออกเป็นส่วนย่อย ๆ ที่มีฐานะเท่าเทียมกัน (Separate-But-Equalists) นักภาษาศาสตร์กลุ่มนี้เชื่อว่าภาษานั้นสามารถแยกออกเป็นส่วนย่อย ๆ ที่มีฐานะเท่าเทียมได้ คือ ด้านการฟัง พูด อ่าน และเขียน แต่ละส่วนมีลักษณะอิสระ แต่มีความสัมพันธ์ซึ่งกันและกัน และมีความสัมพันธ์กับสิ่งแวดล้อมด้วย โดยเชื่อว่าภาษาเป็นพฤติกรรมอย่างหนึ่งของมนุษย์ ที่อาจเรียนรู้ได้ด้วยการตอบสนองต่อสิ่งเร้า นักภาษาศาสตร์พวกนี้มีแนวความเชื่อเกี่ยวกับการเรียนภาษา เช่นเดียวกับพวกนักพฤติกรรมนิยม (Behaviourists) นักภาษาศาสตร์พวกนี้อาจเรียกได้ว่าเป็นพวก โครงสร้างนิยม (structurists) พวกนี้มุ่งศึกษา “การใช้ภาษา” (language usage) นักภาษาศาสตร์กลุ่มนี้นิยมทำการทดสอบความรู้และความสามารถเกี่ยวกับโครงสร้างทางภาษาที่มุ่งศึกษาจากจุดย่อยๆ ที่เชื่อว่ามีผลสำคัญ รวมทั้งทักษะภาษาต่างๆ ด้วยแบบทดสอบปรนัย (objective test) โดยเชื่อว่า หากผู้เรียนสามารถเรียนรู้จุดย่อยๆ ที่สำคัญเหล่านี้ได้แล้วก็สามารถที่จะใช้ภาษาได้ดีในที่สุด
- 3) กลุ่มที่คิดว่าภาษามีสภาพรวมกันอยู่กับสิ่งอื่น ๆ (United Linguists) นักภาษาพวกนี้เชื่อว่าภาษานั้นมีความสัมพันธ์กับสิ่งอื่น ๆ อีกมาก เช่น วัฒนธรรม ฐานะทางสังคม สภาพของจิตใจ และความพร้อม เป็นต้น และเชื่อว่าภาษานั้นไม่สามารถแยกออกเป็นส่วนย่อย ๆ ได้อย่างอิสระ รวมทั้งเชื่อว่าภาษามีการเปลี่ยนแปลงอยู่เสมอ นักภาษาศาสตร์พวกนี้มุ่งศึกษา “ภาษาที่ใช้” (language use) กลุ่มนี้รวมเรียกว่าพวกปฏิวรรต (Transformationists) จากแนวคิดของนักภาษาศาสตร์กลุ่มนี้ทำให้เกิดมีศาสตร์ทางภาษาอีกหลายอย่างเช่น จิตภาษาศาสตร์ (Psycholinguistics) สังคมภาษาศาสตร์ (Sociolinguistic) และ Paralinguistic เป็นต้น ซึ่งศาสตร์ใหม่ ๆ เหล่านี้อาจเรียกรวม ๆ กันว่า “มหภาษาศาสตร์” (Macro-Linguistics) ส่วนศาสตร์ทางภาษาดั้งเดิมนั้นรวมเรียกว่า “จุลภาษาศาสตร์” (Micro-Linguistics) นักภาษาศาสตร์พวกนี้มีแนวคิดเกี่ยวกับทฤษฎีการทดสอบ สิ่งที่ควรทดสอบและวิธีการทดสอบ แตกต่างจากแนวคิดของนักภาษาศาสตร์อีก 2 กลุ่มดังกล่าวแล้วสรุปได้ดังนี้

ตารางที่ 1: ความแตกต่างกันระหว่างทฤษฎีการทดสอบ สิ่งทดสอบ และวิธีการทดสอบของกลุ่มภาษาศาสตร์ต่างๆ

กลุ่ม	ทฤษฎีการทดสอบ	สิ่งที่ทดสอบ	วิธีการทดสอบ
Strict-Separatists	- ทฤษฎีการทดสอบของกลุ่มนี้ "ไม่ได้บ่งบอก" ไว้อย่างชัดเจน แต่เป็นที่เข้าใจกันว่า "ทดสอบเรื่องราวของภาษา" (Test about the language)	- ทดสอบกฎเกณฑ์ต่าง ๆ ทางภาษา โดยมุ่งถึงเรื่องของความถูกต้องตามหลักไวยากรณ์ เช่น การใช้ shall และ will เป็นต้น - ทดสอบการแปล และการวิจารณ์วรรณกรรมที่ได้ศึกษาในชั้น - ทดสอบข้อความที่กำหนดให้ท่องจำ เช่น บทอาขยาน หรือคำศัพท์	- ทดสอบด้วยแบบทดสอบอ้อม เช่น การแต่งประโยค จากคำที่กำหนดมาให้ การแปลข้อความ การวิจารณ์วรรณกรรม หรือการเขียนข้อความที่ให้อ่านเข้าใจ
Separate-But-Equalists	- ผู้นำคนสำคัญทางการทดสอบของกลุ่มนี้คือ Robert Lado (โดยเฉพาะการทดสอบภาษาที่สอง) ผู้ซึ่งเชื่อว่า (Lado, 1977): • ผู้เรียนมีปัญหาในการเรียน ดังนั้นต้องค้นหาปัญหานั้น แล้วทำการทดสอบ • การวิเคราะห์ความแตกต่างมีความจำเป็นสำหรับการค้นหาปัญหาในการเรียน • ฟังทดสอบสิ่งที่สอน และทดสอบเฉพาะปัญหาในการเรียนครั้งละปัญหา • ความตรงร่วมสมัย (concurrent validity) ของแบบทดสอบ เป็นสิ่งจำเป็น • ผู้เรียนไม่สามารถที่จะเรียกได้ว่ารู้ภาษา จนกว่าเขาจะใช้ภาษานั้น ได้อย่างอัตโนมัติและดับไว้ในสถานการณ์ปกติ	- ทดสอบส่วนประกอบทางภาษา 4 อย่างคือ การออกเสียงไวยากรณ์ โครงสร้างและศัพท์ ทั้งในความหมายทางวัฒนธรรมด้วย ทั้งนี้ โดยการทดสอบ 4 ทักษะคือ การฟัง การพูด การอ่าน และการเขียน - ทำการทดสอบส่วนประกอบย่อย ๆ ของส่วนประกอบทางภาษาทั้ง 4 อย่าง จากระดับประโยค จนถึงระดับหน่วยเสียง	- ทดสอบด้วยแบบทดสอบปรนัย เช่น การเลือกตอบ การเติมคำในช่องว่าง การจับคู่ การเรียงส่วนของประโยค และเลือกข้อถูกผิด เป็นต้น - ทดสอบส่วนประกอบของภาษาด้วยข้อทดสอบวัดจุดย่อย (discrete-point items) โดยเน้นหนักที่การทดสอบปัญหาในการเรียนภาษาที่ผู้ได้จากการเปรียบเทียบระหว่างภาษาที่หนึ่งและภาษาที่สอง - ทดสอบด้วยข้อทดสอบจำนวนมากเพราะต้องการให้มีความตรง (validity) และต้องการทดสอบความรวดเร็วในการใช้ภาษาด้วย - นิยมทำการทดสอบด้วยข้อทดสอบแบบเลือกตอบ (multiple-choice tests) ที่มีความเที่ยงตรง (validity) และความเที่ยง (reliability) สูง

ตารางที่ 1: ความแตกต่างกันระหว่างทฤษฎีการทดสอบ สิ่งทดสอบ และวิธีการทดสอบของกลุ่มภาษาศาสตร์ต่างๆ (ต่อ)

กลุ่ม	ทฤษฎีการทดสอบ	สิ่งที่ทดสอบ	วิธีการทดสอบ
United-Linguists	<p>- ผู้นำของกลุ่มนี้เกี่ยวกับการทดสอบทางภาษาคือ Carroll ผู้ซึ่งเชื่อว่าเราต้องตระหนักว่าการแสดงออกทางภาษาของบุคคลมีความเกี่ยวข้องกับความสามารถและความสามารถทางภาษาของตนในลักษณะที่ผสมผสานกัน เช่น การเข้าใจความ" (Carroll, 1972)</p> <p>- นักทดสอบกลุ่มนี้มีความเชื่อว่าภาษานั้นแยกออกเป็นส่วนย่อยๆ ไม่ได้เว้นแต่จะเป็นภาษาในระดับ "จุลภาษา" (micro-language) และในการสื่อความหมายเราต้องใช้ภาษาในระดับ "มหภาษา" (macro-language) ซึ่งต้องใช้ในลักษณะรวมกันหลาย ๆ อย่าง (integrated skills) ดังนั้น การทดสอบก็ต้องมุ่งทำการทดสอบทักษะรวมเหล่านี้พร้อม ๆ กัน เพราะ</p> <ol style="list-style-type: none"> 1. การทดสอบรวมทักษะนี้ ทำให้สามารถทดสอบตัวอย่างของส่วนประกอบทางภาษาได้กว้างขวางยิ่งขึ้น 2. การทดสอบแบบนี้ทำให้การตีความหมายของคะแนนสอบง่ายขึ้น 3. การทดสอบแบบนี้ไม่จำเป็นต้องทำ "การวิเคราะห์ความแตกต่าง (contrastive analysis) ระหว่างภาษาให้เสียเวลา" <p>ทฤษฎีการทดสอบนี้ได้แนวความคิดมาจากความหมายของคำว่าความสามารถ (competence) และความสามารถ (performance)</p>	<p>- ทดสอบความสามารถด้านภาษา (linguistic competence) ในระดับ "จุลภาษา"</p> <p>- ทดสอบความสามารถด้านการสื่อสารนานาชาติ (communicative competence) ในระดับ "มหภาษา"</p> <p>ในทางปฏิบัตินั้น การทดสอบควรคำนึงถึงเรื่อง</p> <ul style="list-style-type: none"> - ปฏิสัมพันธ์ (interaction) - บริบท/เนื้อเรื่อง (context) - ฉาก (setting) - เรื่อง (topic) - หน้าที่ (function) - ทัศนภาวะ (modality) - บทบาท (role) - สถานะเกิดก่อน (pre-supposition) - พิธีการ (formality) - สถานะ (status) - มาลา (mood) - การยอมรับได้ (acceptability) <p>ทางด้าน</p> <ul style="list-style-type: none"> • ภาษา (linguistically) • สังคม (socially) • บริบท (contextually) และ • วัฒนธรรม (culturally) 	<p>- ทำการทดสอบในระดับ "จุลภาษา" ด้วยข้อทดสอบวัดจุดย่อย (discrete-point items) ที่มีลักษณะเป็นปรนัย</p> <p>- ทำการทดสอบในระดับ "มหภาษา" ด้วยข้อทดสอบวัดรวมสรุป (global items) ที่มีลักษณะเป็นปรนัย หรืออัตนัย</p> <p>- ทดสอบทั้งความสามารถด้านความสามารถด้าน</p> <p>- นิยมทำการทดสอบด้วยการเขียนตามคำบอก และแบบทดสอบโคลซ (cloze tests)</p> <p>- นิยมใช้ภาษาตามสภาพจริง (authentic text) ในการทดสอบ</p>

2. ความแตกต่างระหว่างแบบทดสอบจุดย่อยและแบบทดสอบรวมสรุป

แบบทดสอบจุดย่อย (discrete-point test) และแบบทดสอบรวมสรุป (global test/holistic test/integrated test) นั้น แตกต่างกันมากทั้งในด้านรูปแบบและจุดประสงค์ในการใช้ ทั้งนี้เพราะแบบทดสอบชนิดแรกมุ่งทำการทดสอบทางภาษาตามแนวความเชื่อของนักภาษากลุ่มที่ 2 ส่วนแบบที่ 2 นั้นเป็นแนวการทดสอบของนักภาษากลุ่มที่ 3 ซึ่งบางครั้งยังคงอาศัยแบบทดสอบตามแนวแบบทดสอบจุดย่อยมาใช้ด้วย สำหรับการทดสอบความสามารถระยะแรก เพื่อให้เข้าใจความแตกต่างของแบบทดสอบทั้ง 2 ชนิดดังกล่าวแล้ว พร้อมทั้งแนวความเชื่อเบื้องหลัง โปรดศึกษาจากตารางที่ 2 ต่อไปนี้ (Morrow, 1977):

ตารางที่ 2: ความแตกต่างระหว่างแบบทดสอบวัดจุดย่อยและแบบทดสอบวัดรวมสรุป

แบบทดสอบวัดจุดย่อย	แบบทดสอบวัดรวมสรุป
1. เป็นการทดสอบเกี่ยวกับการวิเคราะห์ภาษา (analytical)	1. เป็นการทดสอบเกี่ยวกับการสังเคราะห์ภาษา (synthetical)
2. เป็นการทดสอบปัญหาทีละจุด ๆ (discrete point)	2. เป็นการทดสอบภาษาหลาย ๆ ด้านรวมกัน (holistic)
3. นิยมใช้ประเมินผลด้วยการอิงกลุ่ม (norm-referenced)	3. นิยมใช้ทำประเมินผลด้วยการอิงเกณฑ์ (criterion-referenced)
4. มักใช้ทำการทดสอบด้านการรับรู้ (reception) เช่น การฟังและไวยากรณ์	4. มักใช้ทำการทดสอบด้านการแสดงออก (production) เช่น การพูด และการเขียน
5. ใช้ทดสอบ "ความสามารถศักยภาพทางภาษา" (linguistic competence) เช่น ไวยากรณ์ ระบบการออกเสียง และระบบหน่วยคำ เป็นต้น	5. ใช้ทดสอบ "ความสามารถศักยภาพทางการสื่อสาร" (communicative competence) เช่น การพูด และการเขียน เป็นต้น
6. เป็นการทดสอบระดับ "จุลภาษา" (micro-language)	6. เป็นการทดสอบระดับ "มหภาษา" (macro-language)
7. นิยมใช้เพื่อ "ประเมินผลรวมยอด" (summative evaluation) เช่น การทดสอบใช้แบบทดสอบผลสัมฤทธิ์ (achievement tests)	7. นิยมใช้เพื่อ "ประเมินผลความก้าวหน้า" (formative evaluation) เพื่อรู้ความบกพร่องของการเรียน เช่น การทดสอบใช้แบบทดสอบวินิจฉัย (diagnostic tests)
8. มุ่งทำการทดสอบด้านโครงสร้าง (form) ของภาษา	8. มุ่งทำการทดสอบด้านหน้าที่ (function) ของภาษา
9. เน้นหนักในการทดสอบ "การใช้ภาษา" (language usage)	9. เน้นหนักในการทดสอบ "ภาษาที่ใช้" (language use)
10. การประเมินผลอิงอยู่กับทฤษฎีทางสถิติ (idealization)	10. การประเมินผลอิงอยู่กับคะแนนดิบ (raw data)
11. มักใช้เพื่อทดสอบโครงสร้างลึก (deep structures)	11. มักใช้เพื่อทดสอบโครงสร้างผิว (surface structures)
12. แบบทดสอบคำนึงถึงเรื่องความเที่ยง (reliability) มากกว่าค่าอย่างอื่น	12. แบบทดสอบคำนึงถึงเรื่องความตรง (validity) มากกว่าค่าอย่างอื่น

3. ความแตกต่างของแบบทดสอบสมิทธิภาพทั่วไปกับแบบทดสอบชนิดอื่น

แบบทดสอบมีหลายชนิด ทั้งนี้ขึ้นอยู่กับลักษณะของข้อทดสอบและวัตถุประสงค์ของการทดสอบ เช่น แบบทดสอบผลสัมฤทธิ์ (achievement test) แบบทดสอบสมิทธิภาพ (proficiency test) แบบทดสอบความถนัดเฉพาะด้าน (aptitude test) และแบบทดสอบวินิจฉัย (diagnostic test) เป็นต้น ความแตกต่างของแบบทดสอบเหล่านี้อาจแสดงได้โดยภาพประกอบและตารางต่อไปนี้ (Davies, 1977: 72)

ตารางที่ 3: ความแตกต่างของแบบทดสอบชนิดต่าง ๆ

ชนิดของแบบทดสอบ	เนื้อหา	การสร้าง	การทดสอบ	จุดมุ่งหมาย	ความชัดเจน
แบบทดสอบวัดสมิทธิภาพ	ไม่ชัดเจน	←	X	→	ชัดเจน
แบบทดสอบผลสัมฤทธิ์	ตามหลักสูตร	←	X	→	ชัดเจน/ไม่ชัด
แบบทดสอบความถนัด	ไม่ชัดเจน	←	X	→	ชัดเจน
แบบทดสอบวินิจฉัย	ชัดเจน	←	X	→	ชัดเจน

ตารางที่ 4: ลักษณะเฉพาะที่สำคัญของแบบทดสอบชนิดต่าง ๆ

ที่	ชื่อแบบทดสอบ	กลุ่มข้อทดสอบ	ความยากง่าย	อำนาจจำแนก	หลักสูตร
1	สมิทธิภาพ	วิวิธพันธุ์	ยาก	สูง	ไม่แน่ชัด
2	ผลสัมฤทธิ์	วิวิธพันธุ์	ปานกลาง	ปานกลาง	แน่ชัด
3	ความถนัดเฉพาะ	เอกพันธุ์	ปานกลาง	สูง	ไม่มี
4	วินิจฉัย	เอกพันธุ์	ง่าย	ต่ำ	แน่ชัด

จากข้อมูลในตารางที่ 3 และ 4 ข้างต้น สามารถสรุปได้ว่า

1. **แบบทดสอบสมิทธิภาพ** ได้แก่ แบบทดสอบที่ประกอบด้วยข้อทดสอบที่ไม่ได้ครอบคลุมเนื้อหาของหลักสูตรใดหลักสูตรหนึ่งที่ชัดเจน โดยเฉพาะแต่มีวัตถุประสงค์ที่แน่นอนว่าเพื่อนำผลการทดสอบไปใช้เพื่ออะไรแน่ เช่น เพื่อการศึกษาต่อในระดับที่สูงขึ้น หรือเพื่อการประกอบอาชีพ เป็นต้น นอกจากนี้เนื้อหาของข้อทดสอบจะมีความหลากหลายมากแล้วแต่ที่ผู้สร้างแบบทดสอบจะมีข้อมูลว่ามีเนื้อหาสาระใดที่มีความสำคัญต่อวัตถุประสงค์ที่กำหนดไว้บ้าง จึงทำให้โดยทั่วไปแล้วข้อทดสอบจะยากและมีอำนาจจำแนกสูง แบบทดสอบชนิดนี้ที่รู้จักกันดีทั่วไป ได้แก่ แบบทดสอบ TOEFL (Test of English as a Foreign Language) แบบทดสอบ TOEIC (Test of English as an International Language) และแบบทดสอบเคมบริดจ์ (Cambridge Test) และแบบทดสอบเข้ามหาวิทยาลัย เป็นต้น

2. แบบทดสอบผลสัมฤทธิ์ คือ “เครื่องมือที่สร้างขึ้นเพื่อวัดผลจากการสอนหรือการอบรมรายวิชาใดรายวิชาหนึ่ง” (Anastasi, 1963: 33) ในการใช้แบบทดสอบผลสัมฤทธิ์นั้น “เราสนใจแต่เพียงผลที่เกิดขึ้นภายในกรอบของหลักสูตรที่เราารู้แล้วเท่านั้น” (Davies, 1972: 22) ทั้งนี้ เพื่อทดสอบว่าผู้เรียนเรียนรู้มากน้อยเท่าใด ดังนั้นในการทดสอบทางภาษา เราถือว่า แบบทดสอบผลสัมฤทธิ์นั้นคือ “เครื่องมือที่จะได้มาซึ่งข้อมูลที่บ่งบอกว่า ผู้เรียนสามารถเรียนรู้ได้มากน้อยเท่าใดจากการเรียนรายวิชาใดรายวิชาหนึ่ง” (Clark, 1972:3) ปกติแล้วแบบทดสอบผลสัมฤทธิ์จะประกอบด้วยข้อทดสอบที่มีความหลากหลายเพราะว่าประกอบด้วยเนื้อหากว้างและครอบคลุมทั้งหลักสูตรแต่มีความยาวโดยเฉลี่ยอยู่ในระดับปานกลาง และมีค่าอำนาจจำแนกปานกลาง

ในทางทฤษฎีแล้ว แบบทดสอบผลสัมฤทธิ์และแบบทดสอบวินิจฉัยแตกต่างกันใน 3 ลักษณะคือ

1. หน้าที่
2. เนื้อหา
3. วิธีการวิเคราะห์ข้อทดสอบ

ตารางที่ 5: ความแตกต่างของแบบทดสอบผลสัมฤทธิ์กับแบบทดสอบวินิจฉัย

ลักษณะเฉพาะ	แบบทดสอบผลสัมฤทธิ์	แบบทดสอบวินิจฉัย
1. หน้าที่	<ul style="list-style-type: none"> ● เพื่อใช้ทดสอบว่าผู้เรียนได้เรียนรู้อะไรบ้าง 	<ul style="list-style-type: none"> ● เพื่อใช้ทดสอบว่าผู้เรียนยังไม่ได้เรียนรู้อะไรอีกบ้าง
2. เนื้อหา	<ul style="list-style-type: none"> ● มีเนื้อหาเพื่อการทดสอบอย่างกว้างขวาง 	<ul style="list-style-type: none"> ● มีเนื้อหาเพื่อการทดสอบในวงแคบแต่ใช้จำนวนข้อทดสอบมาก
3. วิธีการวิเคราะห์ข้อทดสอบ	<ul style="list-style-type: none"> ● เน้นหนักในการวิเคราะห์หาความเป็นวิวิธพันธ์ (heterogeneity) ข้อทดสอบ (ข้อทดสอบที่แต่ละข้อมุ่งทดสอบสิ่งที่ต้องการทดสอบต่างกัน) 	<ul style="list-style-type: none"> ● เน้นหนักในการวิเคราะห์หาความเป็นเอกพันธ์ (homogeneity) ของข้อทดสอบ (ข้อทดสอบที่มุ่งทดสอบสิ่งเดียวกัน แต่ถามในหลาย ๆ ด้านต่างกัน)

นอกจากนี้ แบบทดสอบผลสัมฤทธิ์ยังสามารถแบ่งออกได้เป็น 2 ประเภทใหญ่ๆ คือ แบบทดสอบผลสัมฤทธิ์เชิงวินิจฉัย และแบบทดสอบผลสัมฤทธิ์ทั่วไป ซึ่งมีรายละเอียดแตกต่างกันดังนี้ (Clark, 1972)

ตารางที่ 6: ความแตกต่างของแบบทดสอบผลสัมฤทธิ์ 2 ชนิด

ลักษณะเฉพาะ	แบบทดสอบผลสัมฤทธิ์เชิงวินิจฉัย	แบบทดสอบผลสัมฤทธิ์ทั่วไป
1. หน้าที่	<ul style="list-style-type: none"> เพื่อใช้วัดผลสัมฤทธิ์ในการเรียน “จุลภาษา” micro-language ได้แก่ความรู้พื้นฐานเกี่ยวกับระบบหน่วยเสียง หน่วยคำวากยสัมพันธ์ และความสัมพันธ์ระหว่างระบบเหล่านี้ หรือความสามารถศัพทวิทยา (linguistic competence; Rivers, 1972: 33) 	<ul style="list-style-type: none"> เพื่อใช้วัดผลสัมฤทธิ์ในการเรียน “มหภาษา” macro-language ได้แก่การใช้ภาษาเพื่อประโยชน์ ในการสื่อสารนานาชาติ รวมทั้งการให้และการรับ โดยการพูดและการเรียนหรือความสามารถจลนทางภาษารวม (integrated language performance; Rivers, 1972: 33)
2. โครงสร้าง	<ul style="list-style-type: none"> ประกอบด้วย “จุลภาษา” 	<ul style="list-style-type: none"> ประกอบด้วย “มหภาษา”
3. การทดสอบ	<ul style="list-style-type: none"> ทดสอบด้วยข้อทดสอบแบบวัดจุดย่อย เพื่อทดสอบว่าผู้เรียนมี “ผลสัมฤทธิ์” หรือว่า “ไม่มีผลสัมฤทธิ์” ในการเรียน แต่ละจุดย่อยของเนื้อหาที่ต้องการทดสอบ เช่น ทดสอบด้วยแบบทดสอบแบบเลือกตอบ เป็นต้น 	<ul style="list-style-type: none"> ทดสอบด้วยข้อทดสอบแบบวัดรวมสรุป โดยมิได้มุ่งหวังว่าจะทดสอบแต่ละจุดย่อยของเนื้อหา เช่น การทดสอบด้วยการเขียนเรียงความ และแบบทดสอบโคลซ (cloze tests) เป็นต้น

ดังนั้น อาจสรุปได้ว่าแบบทดสอบผลสัมฤทธิ์คือ เครื่องมือที่มุ่งทำการทดสอบว่าผู้สอบหรือผู้เรียนได้เรียนรู้อะไรบ้างในระดับ “จุลภาษา” โดยการทดสอบด้วยแบบทดสอบผลสัมฤทธิ์เชิงวินิจฉัยและได้เรียนรู้อะไรบ้างในระดับ “มหภาษา” โดยการทดสอบด้วยแบบทดสอบผลสัมฤทธิ์ทั่วไป ทั้งนี้การทดสอบทำในกรอบของหลักสูตรที่วางไว้เท่านั้น

- แบบทดสอบความถนัดเฉพาะด้าน ได้แก่แบบทดสอบที่ประกอบด้วยข้อทดสอบที่ครอบคลุมเนื้อหาในหลักสูตรที่มีการเรียนการสอนโดยตรง เช่น ความถนัดทางการจำแนกเสียง ความรอบครอบ และความสามารถในการจำแนกความแตกต่างของภาพที่คล้ายคลึงกัน เป็นต้น ปกติแล้วข้อทดสอบเหล่านี้มีการจัดกลุ่มแยกสิ่งที่ต้องการทดสอบไว้เป็นปัจจัยต่างๆ มีความยากง่ายปานกลาง แต่มีค่าอำนาจจำแนกสูง เช่น แบบทดสอบความถนัดทางภาษา แบบทดสอบความถนัดทางวิศวกรรมศาสตร์ แบบทดสอบแววจู และแบบทดสอบความถนัดทางการเรียน เป็นต้น
- แบบทดสอบวินิจฉัย ได้แก่แบบทดสอบที่มุ่งวิเคราะห์หาจุดบกพร่องและจุดแข็งของผู้เรียนจากการเรียนรู้ในอดีตว่ามีอะไรบ้างเพื่อจะได้ใช้เป็นเครื่องมือในการวางแผนการเรียนซ่อมเสริม (Davies, 1972: 23) แบบทดสอบนี้ประกอบด้วยข้อทดสอบที่ครอบคลุมเนื้อหาของหลักสูตรใดหลักสูตรหนึ่งโดยเฉพาะ หรือ

อาจเป็นเนื้อหาส่วนหนึ่งของหลักสูตรที่ต้องการทดสอบเนื้อหาไม่มากแต่ทดสอบด้วยข้อทดสอบจำนวนมาก ดังนั้น จึงมักเป็นข้อทดสอบที่ง่ายและมีอำนาจจำแนกต่ำ เช่น แบบทดสอบก่อนเริ่มเรียนหลักสูตรใด หลักสูตรหนึ่งที่มีวัตถุประสงค์เพื่อการสำรวจปัญหาในการเรียนหรือแบบทดสอบซ่อมเสริมภายหลังจากการเรียนซ่อมเสริมเนื้อหาบางอย่างของหลักสูตร เป็นต้น

4. ประเด็นปัญหาที่สำคัญของแบบทดสอบสมิทธิภาพทั่วไป

การสร้างแบบทดสอบสมิทธิภาพทั่วไปมีประเด็นปัญหาหลายอย่างที่จะต้องพิจารณา เช่น

1. สมิทธิภาพทั่วไปทางภาษาอังกฤษคืออะไร

คำถามนี้ตอบเฉพาะเจาะจงยากมาก แต่โดยทฤษฎีแล้วสมิทธิภาพทั่วไปทางภาษาอังกฤษก็คือ ความรู้ความสามารถทั่วไปที่จำเป็นสำหรับนิสิต/นักศึกษาต้องมีเพื่อใช้เป็นเครื่องมือในการศึกษาต่อในระดับมหาวิทยาลัย ทั้งในประเทศและต่างประเทศ และความรู้ความสามารถดังกล่าวนี้ได้มากจากการเลือกสรรของผู้สอนที่มีประสบการณ์ในการเรียนการสอนในระดับมหาวิทยาลัยมาแล้วเป็นอย่างดี

2. เราสามารถจะวัดสมิทธิภาพทั่วไปทางภาษาได้หรือไม่

คำถามนี้ตอบยาก เพราะมีปัจจัยหลายอย่างที่เกี่ยวข้องจนกลายเป็นปัญหาขัดเคืองในการทดสอบสมิทธิภาพทั่วไปทางภาษา ก่อนที่จะตอบคำถามนี้เราควรทำความเข้าใจกับปัญหาขัดเคืองดังต่อไปนี้คือ

ก. ปัญหาเกี่ยวกับความตรงในการทดสอบ

ในการทดสอบทางภาษานั้นเป็นการยากมากที่จะรู้ว่า สิ่งที่เราทดสอบนั้นคือสิ่งที่เราต้องการทดสอบอย่างแท้จริงหรือไม่ หรือพูดอีกประการหนึ่งก็คือ เป็นการยากที่จะรู้ว่าข้อทดสอบของเรานั้นมีความตรง (validity) หรือไม่ แม้ว่าแบบทดสอบจะมีความเที่ยง (reliability) สูงมากเพียงใดก็ตาม (เพราะว่าความเที่ยงสามารถเพิ่มหรือลดได้โดยวิธีสถิติ) การหาความตรงของข้อทดสอบนั้น เราไม่มีวิธีการที่จะหาได้อย่างแท้จริงโดยวิธีการทางวิทยาศาสตร์ แต่เราหาได้แต่เพียงการอ้างอิงเกณฑ์อื่น หรืออาศัยตรรกวิทยาเท่านั้น (Carroll, 1972: 8) สิ่งที่เรารู้แน่ไม่ได้ก็คือสิ่งที่เราทดสอบนั้น คือ สิ่งที่เราต้องการทดสอบหรือไม่ เพราะว่า เป็นการยากที่จะรู้ว่าสมิทธิภาพทั่วไปในการเรียนภาษาแต่ละด้านนั้นประกอบด้วยอะไรบ้าง โดยเฉพาะอย่างยิ่งการทดสอบในระดับ “มหภาษา” เช่น การอ่านเข้าใจความ และการฟังเข้าใจความ (reading comprehension and listening comprehension) เป็นต้น เพราะว่า เป็นการยากที่จะรู้ว่า “การเข้าใจความ” นั้น ประกอบด้วย ส่วนประกอบทางภาษาอะไรบ้าง นักทดสอบทางภาษาก็มีความเห็นแตกต่างกันในเรื่องส่วนประกอบของภาษาในแต่ละด้าน ยกตัวอย่างเช่น Davies และ Widdowson (Davies and Widdowson, 1974: 168-176) เชื่อว่าการอ่านเข้าใจความสามารถวัดได้จากด้านต่าง ๆ ต่อไปนี้คือ

- ก. การอ้างอิงโดยตรง (direct reference)
- ข. การอนุมานความหมาย (inference)
- ค. การสมมุติ (supposition)
- ง. การประเมินผล (evaluation)

แต่ว่าการทดสอบการอ่านเข้าใจความของ TOEFL (Test of English as a Foreign Language) นั้นจะทำการทดสอบเกี่ยวกับความรู้และความสามารถในการเรื่องดังต่อไปนี้ด้วย คือ

- ก. การตีความหมายของเนื้อหา (paraphrasing)
- ข. การทดสอบศัพท์ในและนอกเนื้อเรื่อง (vocabulary testing in and out of the context) และ
- ค. การสรุปผลของเนื้อเรื่อง (conclusion)

ดังนั้น เป็นการยากที่จะกล่าวได้ว่า สิ่งที่เราทดสอบนั้นคือสิ่งที่เราต้องการทดสอบอย่างแท้จริงหรือไม่ เพราะเป็นการยากยิ่งที่จะรู้ส่วนประกอบที่แท้จริงของเนื้อหาบางอย่างทางภาษา โดยเฉพาะอย่างยิ่งการทดสอบระดับ “มหากาษา” แม้ว่าเราจะหาความตรงของแบบทดสอบได้หลายวิธีก็ตาม (Cronbach, 1960: 106) เช่น

1. ความตรงเชิงเนื้อหา (content validity) โดยการเปรียบเทียบข้อทดสอบกับเนื้อหาที่สำคัญที่มุ่งทดสอบ
2. ความตรงร่วมสมัย (concurrent validity) โดยการหาความสัมพันธ์ระหว่างผลของการทดสอบกับผลของการทดสอบของแบบทดสอบที่ดีที่มีอยู่แล้ว
3. ความตรงเชิงพยากรณ์ (predictive validity) โดยการหาความสัมพันธ์ระหว่างผลของการทดสอบนั้น กับผลิตผลสุดท้ายที่ใช้เป็นเกณฑ์ในอนาคต
4. ความตรงเชิงภาวะสันนิษฐาน (construct validity) โดยการเปรียบเทียบผลของการทดสอบกับแนวความเชื่อหรือทฤษฎีที่ตั้งขึ้น

แต่ว่าการหาความตรงโดยวิธีทางสถิตินี้ก็ไม่ใช่หลักประกันว่า สิ่งที่เราทดสอบนั้นคือสิ่งที่เราต้องการทดสอบอย่างแท้จริง แม้ว่าจะเป็นแนวทางที่บ่งให้รู้ถึงแนวโน้มว่าสิ่งที่เราต้องการทดสอบนั้นคืออะไร (Carroll, 1972: 9)

ดังนั้น เราจึงสามารถสรุปได้ว่า การหาความตรงของสิ่งที่เราต้องการวัดนั้นอาจสามารถทำได้ในการทดสอบภาษาระดับ “จุลภาษา” ส่วนในระดับ “มหากาษา” นั้นทำได้ยาก โดยเฉพาะอย่างยิ่งการทดสอบ “ความสามารถศักย์ในการสื่อสาร” (communicative competence) เช่น ความสามารถในการใช้ภาษาต่างประเทศโดยการเลือกใช้รูปแบบของภาษา (registers) และสัญชาติญาณที่ใกล้เคียงกับเจ้าของภาษาได้อย่างเหมาะสมและถูกต้องตามความหมายของเนื้อหาที่สื่อสารในสถานการณ์ที่เป็นจริง (Jakobovits, 1970)

ข. ปัญหาเกี่ยวกับขอบข่ายของเนื้อหา

นอกจากนี้ปัญหาในการทดสอบสมรรถภาพทั่วไปทางภาษาก็คือ เราไม่ทราบแน่ชัดว่า “เราจะต้องทำการทดสอบเนื้อหาในการสอนนั้นมากน้อยสักเพียงใด” จึงจะครอบคลุมกับความคิดรวบยอดเกี่ยวกับความสามารถอย่างแท้จริงในการใช้ภาษาต่างประเทศ “และเราก็ไม่รู้แน่ว่าภาษานั้นสามารถที่จะแยกออกเป็น ส่วนประกอบย่อย ๆ เพื่อการทดสอบได้หรือไม่ (Carroll, 1972: 11) ที่เป็นเช่นนี้ก็เพราะว่า ค่าสหสัมพันธ์ของผลการทดสอบทักษะต่าง ๆ อาจจะแตกต่างกันได้แล้วแต่ว่าการสอนและการทดสอบจะเน้นหนักไปในแนวทางใด ยกตัวอย่างเช่น ค่าสหสัมพันธ์ระหว่างการอ่านเข้าใจความและการฟังเข้าใจความมีสูง ไม่ได้หมายความว่า ผู้เรียนมีผลสัมฤทธิ์ในการเรียนทักษะทั้งสองเท่า ๆ กัน (เพราะค่าเฉลี่ยของการทดสอบนั้นอาจแตกต่างกันอย่างมีนัยสำคัญก็ได้)

นอกจากนี้ “ความสามารถในการใช้ภาษาต่างประเทศ ไม่ใช่ผลผลิตรวมของส่วนย่อยต่าง ๆ ทางภาษา” (Rivers, 1972: 32) ทั้งนี้ก็เพราะว่า ผู้ที่มีความสามารถในระดับ “มหภาษา” มาก สามารถที่จะมีผลสัมฤทธิ์ในระดับ “จุลภาษา” มาก ได้ด้วย แต่ผู้ที่มีความสามารถในระดับ “จุลภาษา” มาก อาจจะไม่มีผลสัมฤทธิ์ในระดับ “มหภาษา” ในระดับที่ยอมรับได้ก็เป็นได้

ค. ปัญหาเกี่ยวกับประสิทธิภาพในการทดสอบ

ปัญหาอีกประการหนึ่งก็คือ ปัญหาเกี่ยวกับการทำการทดสอบ เช่น การเก็บรักษาข้อทดสอบไว้เป็นความลับ ค่าของเวลาและพลังงานในการสร้างข้อทดสอบ การทำการทดสอบ การให้คะแนน และการให้ระดับคะแนน เป็นต้น สิ่งต่าง ๆ เหล่านี้ล้วนแต่มีผลโดยตรงกับความตรง ความเที่ยง และประสิทธิภาพของแบบทดสอบด้วย ยกตัวอย่างเช่น “การที่จะทำให้แบบทดสอบมีความตรง และความเที่ยงมากขึ้น เวลาในการทดสอบต้องมีเพียงพอ และควรต้องใช้โสตทัศนูปกรณ์ช่วยด้วย” (Carroll, 1972: 14)

ง. ปัญหาเกี่ยวกับการสอนเพื่อการทดสอบ

มีบ่อยครั้งที่ผู้สอน “สอนตามข้อทดสอบ” ทำให้ผลที่ได้คือความไม่สมบูรณ์ของการทดสอบ เพราะว่าการสอนเช่นนี้มักเน้นที่เทคนิคการทดสอบข้อทดสอบมากกว่าเนื้อหาที่ต้องการทดสอบ ปัญหาเช่นนี้เกิดขึ้นได้ง่าย โดยเฉพาะอย่างยิ่งการสอนเพื่อสอบข้อทดสอบที่ผู้อื่นสร้างขึ้น

จ. ปัญหาเกี่ยวกับการประยุกต์ใช้ทฤษฎีการทดสอบและวัดผล

มีบ่อยครั้งเช่นกันที่ทฤษฎีเกี่ยวกับการทดสอบและวัดผลยากแก่การนำมาประยุกต์ใช้ได้ อาจจะเป็นเพราะผู้สร้างข้อทดสอบขาดความรู้ในทฤษฎี หรือขาดความชำนาญในการประยุกต์ใช้ทฤษฎี นอกจากนี้ปัญหานี้ อาจเกี่ยวข้องกับเวลาและค่าใช้จ่ายในการพัฒนาแบบทดสอบด้วย เลยทำให้ยากแก่การนำมาประยุกต์ใช้ เช่น การสัมภาษณ์เพื่อทดสอบการฟังและการพูดผู้เรียนจำนวนมากทำให้เสียเวลามาก ถ้าหากว่าจะใช้เครื่อง

บันทึกเสียงในห้องปฏิบัติการภาษาก็ราคาแพง และผู้สอนบางคนไม่มีความรู้ในการควบคุมเครื่องหรือการวิเคราะห์ข้อทดสอบตามทฤษฎีการตอบสนองต่อข้อทดสอบ (Item Response Theory) ก็เข้าใจยาก เป็นต้น

ดังนั้น เราอาจสรุปได้ว่า การทดสอบสมรรถภาพทั่วไปทางภาษานั้นยากมาก แต่ก็ไม่ใช่ว่าจะทำไม่ได้ วิธีหนึ่งที่เราจะทำได้ในขณะนี้ก็คือการทดสอบความรู้และความสามารถทั่วไปโดยแบบทดสอบที่ประกอบด้วยข้อทดสอบวัดความสามารถ (performance) และความสามารถศักย์ (competence) ทางภาษา เช่น ทดสอบการอ่านเข้าใจความด้วยข้อทดสอบผลผลิตทางภาษา (language production: เช่น การเขียน และการพูด เป็นต้น) ด้วยการวัดความสามารถในการใช้ภาษาตามสภาพความเป็นจริง เป็นต้น

นอกจากนี้ เนื้อหาทางภาษาที่ทำการทดสอบจะต้องมีความเหมาะสมกับสภาพความเป็นจริงด้วย กล่าวคือ ผู้สอนจะต้องคำนึงว่าภาษาที่ใช้นั้นเป็นภาษาที่ทำให้ง่ายขึ้น (simplified text) หรือว่าเป็นภาษาที่ทำให้ยากขึ้น (complexified text) หรือว่าเป็นภาษาที่ใช่จริง (authentic text) ภาษาที่ใช้ประกอบด้วย ส่วนประกอบทางภาษา (linguistic components) ที่ใช้กันอยู่เสมอ หรือว่าเป็นภาษาที่ไม่ค่อยมีคนนิยมใช้ ในการทดสอบนั้นควรจะเน้นหนักไปในทางส่วนประกอบใด และจะต้องตระหนักเสมอว่าการเรียนรู้ภาษานั้น จะต้องคำนึงถึงความสามารถด้านความรวดเร็วและความถูกต้องด้วยไม่ใช่แต่เพียงความรู้และความสามารถในเชิงเนื้อหาเท่านั้น เป็นที่น่าสังเกตว่าแนวโน้มในการทดสอบในปัจจุบันนี้ มุ่งทำการสอนและสอบ “ภาษาที่ใช่จริง” ที่ประกอบด้วยส่วนประกอบทางภาษา (linguistic components) ที่นิยมใช้กันอยู่อย่างแพร่หลาย แทนที่จะเป็นเนื้อหาจาก “ภาษาที่ทำให้ง่ายขึ้น” หรือ “ภาษาที่ทำให้ยากขึ้น”

5. ลักษณะที่ดีของแบบทดสอบสมรรถภาพทั่วไปทางภาษา

แบบทดสอบสมรรถภาพทั่วไปทางภาษาที่ดีควรมีลักษณะเหมือนกับแบบทดสอบที่ดีทั่วไป 3 ประการ (Marshall and Hales, 1972: 4; Ebel, 1972: 359, Hambleton 1980: 209) คือ

1. มีความตรง (validity) มาก หมายความว่าข้อทดสอบแต่ละข้อทดสอบสามารถวัดสิ่งที่ต้องการวัดอย่างแท้จริงได้มาก ปัจจัยสำคัญที่จะทำให้เกิดความตรงที่สำคัญคือ
 - 1) ความสอดคล้องระหว่างข้อทดสอบกับวัตถุประสงค์ในการทดสอบ (item-objective congruence) ว่ามีมากน้อยเพียงใด ผู้พัฒนาแบบทดสอบอาจทำการทดสอบค่าดัชนีความสอดคล้อง (item-congruence index) เป็นต้น
 - 2) ความยุติธรรมหรือไม่มีอคติต่อสิ่งที่ทดสอบ (fairness or unbiased) คือการที่ข้อทดสอบให้ความยุติธรรมแก่ผู้สอบทุกคนเสมอหน้ากัน ข้อทดสอบไม่เอื้อผลประโยชน์หรือลำเอียงเข้าข้างผู้สอบกลุ่มใดกลุ่มหนึ่ง โดยเฉพาะ ข้อทดสอบที่มีอคติ (biased item) จะทดสอบสิ่งที่ไม่ได้ตั้งใจจะทดสอบ ซึ่งอาจจะเป็นเพราะ (Shepard, 1982: 10)

- ก. เกิดจากการเลือกเนื้อหาอย่างมีอคติ (bias in selection) เช่น ผู้สร้างข้อทดสอบเลือกเนื้อหาบางส่วนเท่านั้นทำให้เนื้อหาไม่ครอบคลุมสิ่งที่ต้องการทดสอบทั้งหมดและทำให้แบบทดสอบไม่มีอัตราส่วนของเนื้อหาที่สมดุลกัน
- ข. เกิดจากการเขียนข้อทดสอบ (bias in test construction) คือการที่ผู้สร้างข้อทดสอบเขียนข้อความให้เกิดประโยชน์แก่คนบางคนหรือบางกลุ่ม เช่น ใช้คำศัพท์ที่ใช้ในกลุ่มผู้สอบกลุ่มใดกลุ่มหนึ่งโดยเฉพาะจึงทำให้ผู้ใช้ภาษาทั่วไปเสียเปรียบ เป็นต้น

ผู้สร้างแบบทดสอบอาจจะตรวจสอบได้เองว่าข้อทดสอบมีความยุติธรรมหรือไม่โดยการพิจารณาจากภาษาที่ใช้และวิธีการเลือกเนื้อหาทำการทดสอบ หรืออาจทำการทดสอบทางสถิติก็ได้ (Rudner and Others 1980: 9)

- 3) ความเป็นปรนัยของข้อทดสอบ (item objectivity) ได้แก่ความถูกต้องและคงที่ของคำตอบของข้อทดสอบแต่ละข้อ ไม่ว่าผู้ตรวจคำตอบจะเป็นผู้ใดก็ตามจะต้องได้ผลคงที่เสมอ ซึ่งหมายความว่าแบบทดสอบที่ดีจะต้องมีคำตอบที่คงที่แน่นอน
2. **มีความเที่ยง (reliability) สูง** หมายความว่าแบบทดสอบสามารถให้ผลการทดสอบคงที่เสมอ ซึ่งปัจจัยที่ทำให้ได้ผลดังกล่าวนี้มีหลายอย่าง เช่น
- 1) ค่าอำนาจจำแนกของข้อทดสอบ (item discrimination) ของแบบทดสอบสมมติภาพทั่วไปสำหรับแบบทดสอบอิงกลุ่ม (norm-referenced test) ที่ดีควรมีค่าอำนาจจำแนกตั้งแต่ $r_{pb} > 0.20$ ขึ้นไป แต่ทั้งนี้ แล้วแต่ลักษณะของกลุ่มของผู้เข้ารับการทดสอบ กล่าวคือ หากว่าเป็นกลุ่มเอกพันธ์ (homogeneous group) ค่าอำนาจจำแนกอาจต่ำหรือต่ำมาก แต่หากเป็นกลุ่มวิวิธพันธุ์ (heterogeneous group) ค่าอำนาจจำแนกควรสูงกว่าค่าดังกล่าวมาก เช่น $r_{pb} > 0.30$ เป็นต้น
 - 2) ความยาวที่เหมาะสม (appropriate length) แบบทดสอบที่ดีควรมีความยาวสูงสุดไม่เกิน 300 ข้อ และ/หรือ ใช้เวลาทดสอบไม่เกิน 3 ชั่วโมง เพราะแบบทดสอบที่ยาวเกินไปทำให้มีผลเสียทางอารมณ์และทำให้เกิดการเตาในการทดสอบ
 - 3) ความเฉพาะของข้อทดสอบ (item specification) ได้แก่การที่ข้อทดสอบถามเฉพาะเจาะจงสิ่งที่ต้องการทดสอบได้สอดคล้องกับวัตถุประสงค์ของการทดสอบแต่ละข้อของเนื้อหาที่ต้องการทดสอบ เช่น ระดับความรู้ ความจำ การวิเคราะห์ การสังเคราะห์ และการประเมินผล เป็นต้น
 - 4) ความเหมาะสมของเกณฑ์การตัดสิน (appropriate criteria) วิธีการตัดสินผลการทดสอบมีผลต่อความเที่ยงของแบบทดสอบ ผู้ใช้ผลการทดสอบควรต้องคำนึงความคลาดเคลื่อนมาตรฐานในการวัด (standard error of measurement) ด้วย โดยเฉพาะเมื่อต้องตัดสินใจผลการทดสอบ

5) ความยากของข้อทดสอบ (item difficulty) แบบทดสอบสมิทริภาพทั่วไปสำหรับแบบทดสอบอิงกลุ่มควรประกอบด้วยข้อทดสอบที่มีระดับความยากระหว่าง $p = 0.15-0.85$ และมีค่าเฉลี่ยของความยากประมาณ 0.40-0.60 ค่าความยากของข้อทดสอบควรจะมีการกระจายเป็นรูปโค้งปกติ (normal curve) และควรจะเรียงข้อทดสอบตามระดับความยากของข้อทดสอบจากข้อที่ง่ายที่สุดไปหาข้อที่ยากที่สุด เพราะว่าความยากของข้อทดสอบมีผลต่อการทดสอบโดยเฉพาะความเที่ยงอย่างมีนัยสำคัญ (Plake and Others, 1982: 55)

อนึ่ง การที่จะรู้ระดับความยากและอำนาจจำแนกของข้อทดสอบจำเป็นต้องมีการวิเคราะห์ข้อทดสอบรายข้อ (Item analysis) ตามทฤษฎีการตอบสนองต่อข้อทดสอบ (Item Response Theory) เช่น การวิเคราะห์ข้อทดสอบตามวิธีรูปแบบราซ (Rasch Model) หรือรูปแบบ 3 พารามิเตอร์ (Three-Parameter Model) เป็นต้น ซึ่งเป็นขั้นตอนหนึ่งที่สำคัญในการพัฒนาแบบทดสอบ

3. มีความสะดวกในการใช้ (practicality) ได้แก่ ความสะดวกสบายรวมทั้งความประหยัดในการใช้แบบทดสอบนั้น ซึ่งอาจพิจารณาได้จาก

- 1) สะดวกในการตรวจ (scorability)
- 2) สะดวกในการบริหารการทดสอบ (administrability)

เมื่อเราทราบว่าแบบทดสอบสมิทริภาพทั่วไปที่ดีเป็นอย่างไร ขึ้นต่อไปที่สำคัญมากก็คือเราจะมีวิธีการสร้างและพัฒนาแบบทดสอบได้อย่างไร

ส่วนที่ 2 การสร้างและการพัฒนาแบบทดสอบสมิทริภาพทั่วไปทางภาษา

1. ขั้นตอนในการสร้างและพัฒนาแบบทดสอบสมิทริภาพทั่วไป

ในการสร้างและพัฒนาแบบทดสอบสมิทริภาพทั่วไปนั้นมีขั้นตอนที่สำคัญ 8 ขั้นตอน เช่นเดียวกับการสร้างแบบทดสอบอื่นทั่วไป ดังนี้

เป้าหมาย:

ตัดสินผลการทดสอบว่าผู้สอบ
มีความรู้ความสามารถอยู่ในระดับใด
เพื่อประโยชน์ในการใช้งานหรือสอนต่อไป



8. นำผลการทดสอบไปใช้
7. ประเมินผลแบบทดสอบ
6. นำแบบทดสอบไปใช้
5. รวบรวมข้อทดสอบให้เป็นแบบทดสอบ
4. เตรียมข้อทดสอบที่เหมาะสม
3. เลือกชนิดของข้อทดสอบที่เหมาะสม
2. สร้างตารางกำหนดเนื้อหาแบบทดสอบ
1. กำหนดวัตถุประสงค์ของการทดสอบ

ภาพประกอบที่ 1: ขั้นตอนการสร้างและพัฒนาแบบทดสอบ

แต่ละขั้นตอนมีรายละเอียดที่สำคัญดังต่อไปนี้

ขั้นที่ 1 กำหนดวัตถุประสงค์ของการทดสอบ (Determining the Purposes of Testing)

ปกติแล้วแบบทดสอบสมิทธิภาพใช้เพื่อทดสอบผู้เรียนว่ามีความรู้ความสามารถ ทัว ๆ ไป มักใช้เพื่อการวัดผลการรวบยอด (summative evaluation) เนื้อหาต่าง ๆ ที่ผู้เรียนเคยเรียนมาแล้วหลาย ๆ อย่างหรือหลาย ๆ วิชา

อนึ่ง ในการกำหนดวัตถุประสงค์ของการทดสอบนั้นผู้สร้างแบบทดสอบจำเป็นจะต้องกำหนดวัตถุประสงค์ย่อย (sub-objectives) ของการทดสอบด้วยว่าต้องการจะทำการทดสอบอะไร ซึ่งปกติแล้วนักทดสอบนิยมกำหนดเป็นวัตถุประสงค์เชิงพฤติกรรม (behavioral objective) ซึ่งมีคุณลักษณะเฉพาะที่สำคัญ 3 ประการคือ

1. มีพฤติกรรมเป้าหมาย (target behavior) ที่เฉพาะเจาะจงและแจ่มชัด
2. มีมาตรฐานการปฏิบัติ (standard performance) อย่างชัดเจนเป็นเกณฑ์
3. มีเงื่อนไข (condition) หรือสถานการณ์ (situation) กำกับ

ตัวอย่างเช่น เมื่ออ่านเรื่องที่กำหนดให้แล้ว ผู้สอบสามารถตอบข้อทดสอบได้ถูกต้องอย่างน้อย 8 ข้อ จาก 10 ข้อ ที่ถามเกี่ยวกับความเข้าใจเรื่องที่อ่าน เป็นต้น และในการทดสอบตามระดับชั้นของความรู้หรืออนุกรมวิธาน (taxonomy) ผู้สร้างข้อทดสอบอาจศึกษาได้จากหลายแหล่ง เช่น อนุกรมวิธานของ Bloom (Bloom 1968) ของ Barrette (Clymer, 1983: 56) และของ Valette and Disick (Valette and Disick, 1972) เป็นต้น สำหรับทางการศึกษาทั่วไปนิยมใช้แนวคิดของ Bloom เป็นแนวทางในการกำหนดวัตถุประสงค์ในการเรียนการสอนและการทดสอบ ซึ่งแนวคิดนี้แบ่งระดับชั้นของความรู้ทางด้านพุทธิพิสัย (cognitive domain) จากง่ายที่สุดไปหายากที่สุด 6 ชั้น (Bloom, 1968: 271-273) คือ

1. ความรู้ (knowledge)

1.1 ความรู้เฉพาะ (knowledge of specifics)

1.1.1 ความรู้ในศัพท์เฉพาะ (knowledge of terminology)

1.1.2 ความรู้เกี่ยวกับความจริงเฉพาะ (knowledge of specific facts)

1.2 ความรู้เกี่ยวกับวิธีการจัดการกับข้อเท็จจริงเฉพาะ (knowledge of way of dealing with specifics)

1.2.1 ความรู้เกี่ยวกับแบบแผนนิยม (knowledge of conventions)

1.2.2 ความรู้เกี่ยวกับแนวโน้มและลำดับขั้นตอน (knowledge of classification and categories)

1.2.3 ความรู้เกี่ยวกับการจำแนกและจัดประเภท (knowledge of classification and categories)

1.2.4 ความรู้เกี่ยวกับเกณฑ์ (knowledge of criteria)

1.2.5 ความรู้เกี่ยวกับวิธีดำเนินการ (knowledge of methodology)

1.3 ความรู้ในเรื่องของหลักการทั่วไปและความเป็นนามธรรมในสาขา (knowledge of universals and abstractions in a field)

1.3.1 ความรู้เกี่ยวกับหลักการและข้อสรุปทั่วไป (knowledge of principles and generalizations)

1.3.2 ความรู้เกี่ยวกับทฤษฎีและโครงสร้าง (knowledge of theories and structures)

2. ความเข้าใจ (comprehension)

2.1 การแปล (translation)

2.2 การตีความ (interpretation)

2.3 การขยายความ (extrapolation)

3. การประยุกต์ (application)

4. การวิเคราะห์ (analysis)

- 4.1 การวิเคราะห์ส่วนประกอบ (analysis of elements)
- 4.2 การวิเคราะห์ความสัมพันธ์ (analysis of relationships)
- 4.3 การวิเคราะห์หลักการการจัดระบบ (analysis of organization principles)

5. การสังเคราะห์ (synthesis)

- 5.1 การผลิตสิ่งสื่อความหมายเฉพาะ ((production of a unique communication)
- 5.2 การจัดวางแผน โครงการหรือการเสนอแผนการดำเนินงาน (production of a plan or proposed set of operations)
- 5.3 การค้นหาความสัมพันธ์เชิงนามธรรม (derivation of a set of abstract relations)

6. การประเมินค่า (evaluation)

- 6.1 การตัดสินคุณค่าโดยใช้เกณฑ์ภายใน (judgment in term of internal evidence)
- 6.2 การตัดสินคุณค่าโดยใช้เกณฑ์ภายนอก (judgment in term of external criteria)

ดังนั้นผู้สร้างแบบทดสอบจะต้องกำหนดว่าแบบทดสอบที่จะสร้างขึ้นนั้นต้องการทดสอบความรู้หรือความสามารถด้านใดของผู้เรียนโดยเฉพาะ และวัตถุประสงค์เฉพาะนี้จะต้องสอดคล้องกับวัตถุประสงค์ของการทดสอบที่ผู้สร้างแบบทดสอบกำหนดไว้ด้วย

อนึ่ง สำหรับการทดสอบความรู้ความสามารถทางภาษา ผู้สร้างข้อทดสอบอาจอาศัยหรือประยุกต์แนวคิดของอนุวิธานของ Bloom หรือของผู้อื่น เช่น Valette and Disick มาใช้ก็ได้

ขั้นที่ 2 สร้างตารางกำหนดเนื้อหาแบบทดสอบ (Building a Table of Specification)

ตารางกำหนดเนื้อหาของแบบทดสอบเปรียบเสมือนพิมพ์เขียว (blue print) ของการสร้างสิ่งก่อสร้างต่าง ๆ ดังนั้น จึงมีความสำคัญมาก เพราะตารางนี้จะบอกให้ผู้สร้างแบบทดสอบรู้ว่าเนื้อหาของแบบทดสอบจะสอดคล้องกับสิ่งที่ต้องการหรือไม่ และการกำหนดเนื้อหาของแบบทดสอบจะเป็นอัตราส่วนที่เหมาะสมกับเนื้อหาที่ต้องการ ตารางกำหนดเนื้อหามีลักษณะเป็นตาราง 2 มิติที่แสดงความสัมพันธ์ของเนื้อหาของสิ่งที่ต้องการทดสอบกับเนื้อหาของข้อทดสอบ และกำหนดน้ำหนักของสำคัญของเนื้อหา ซึ่งมีลักษณะดังนี้

ตารางที่ 7: ตัวอย่างตารางกำหนดเนื้อหาของแบบทดสอบสมรรถภาพทั่วไป
ด้านการอ่านเข้าใจความ (เป็นร้อยละ)

Objectives						
Passage	Knowledge	Comprehension	Application	Analysis/ Synthesis	Evaluation	Total
Food	6	6	8	8		28
Transportation	6	6	4	8		24
City Life	4	4	8	5	5	26
Farm Life	4	4	5	4	5	22
Total	20	20	25	25	10	100

ข้อควรสังเกตก็คือตัวเลขในตารางแสดงเป็นร้อยละของจำนวนข้อทดสอบทั้งหมด 100 ข้อ ซึ่งในทางปฏิบัตินั้นผู้สร้างแบบทดสอบอาจกำหนดน้ำหนักของเนื้อหาของแบบทดสอบเป็นร้อยละหรือจำนวนข้อด้วยก็ได้ เพื่อให้ง่ายต่อการสร้างข้อทดสอบ แต่นักทดสอบมักจะสร้างข้อทดสอบไว้เพื่อเลือกอีกประมาณร้อยละ 10-20 แล้วแต่ความชำนาญในการสร้างมีมากน้อยเพียงใด

ส่วนการกำหนดน้ำหนักของเนื้อหาหรือจำนวนข้อทดสอบนั้น กำหนดจากความสำคัญของเนื้อหาต่อจำนวนเวลา (นาที) ทั้งหมดที่ใช้ในการทดสอบโดย

1. จัดทำรายการเนื้อหาที่สำคัญที่จะทำการทดสอบทั้งหมด
2. ระบุวัตถุประสงค์ย่อยเฉพาะของการทดสอบทั้งหมด
3. กำหนดความยาวของแบบทดสอบว่าเป็นเท่าใด (กี่ข้อ) จึงจะเหมาะสมกับเวลาในการทดสอบ
4. กำหนดน้ำหนักแต่ละเนื้อหาที่จำเป็นต่อเป้าหมายของการทดสอบว่าเป็นร้อยละเท่าใด ตามความสำคัญของเนื้อหา เช่น เนื้อหาทั้งหมดมี 25 หัวข้อ ดังนั้น เมื่อต้องการสร้างข้อทดสอบ 100 ข้อ ก็ย่อมจะต้องสร้างข้อทดสอบวัดเนื้อหาละประมาณ 4 ข้อ และอาจเผื่อไว้อีก 1-2 ข้อ (แล้วจะได้ตัวเลขในแถวสุดท้าย)
5. พิจารณาปรับความสำคัญของเนื้อหาที่จำเป็นต่อเป้าหมายของการทดสอบนั้นว่าควรเน้นหนักที่วัตถุประสงค์ใดมากที่สุดหรือน้อยกว่ากัน และคิดเป็นร้อยละเท่าใด (แล้วจะได้ตัวเลขในแถวสุดท้าย)

6. พิจารณาแบ่งร้อยละของความสำคัญทั้งในแนวเนื้อหาและวัตถุประสงค์ออกเป็นร้อยละย่อยให้สอดคล้องกับวัตถุประสงค์ของการทดสอบ

7. ตรวจสอบความสอดคล้องของร้อยละต่าง ๆ ทั้ง 2 มิติให้ถูกต้อง

อนึ่ง มีนักทดสอบบางท่านนิยามกำหนดน้ำหนักของเนื้อหาของแบบทดสอบสมิทธิภาพทั่วไปโดยอาศัยจำนวนวัตถุประสงค์เชิงพฤติกรรมที่ต้องการทดสอบว่าข้อใดมีความสำคัญมากน้อยกว่ากันแล้วจึงกำหนดความยาวที่เหมาะสมของเวลาสอบในภายหลัง

ขั้นที่ 3 เลือกชนิดของข้อทดสอบที่เหมาะสม (Selecting Appropriate Item Types)

ข้อทดสอบมีหลายชนิดแตกต่างกัน เช่น

1. แบบเขียนตอบ (supply type) ซึ่งแบ่งเป็น

1.1 แบบอัตนัยไม่จำกัดคำตอบ (essay-extended response)

1.2 แบบอัตนัยจำกัดคำตอบ (essay-restricted response)

1.3 แบบทดสอบโคลซ (cloze) ซึ่งมีหลายรูปแบบ

1.4 แบบตอบสั้น ๆ (short answer)

1.5 แบบเติมความให้สมบูรณ์ (completion)

2. แบบเลือกตอบ (selecting type) ซึ่งอาจแบ่งเป็น

2.1 แบบถูก-ผิด (true-false)

2.2 แบบจับคู่ (matching)

2.3 แบบเลือกตอบหลายตัวเลือก (multiple-choice)

2.4 แบบเลือกตอบหลายตัวเลือกและค่าน้ำหนักคะแนนต่างกัน (multiple-point and multiple-choice)

ผู้สร้างแบบทดสอบจะต้องพิจารณาว่าจะเลือกข้อทดสอบแบบใดจึงจะสามารถวัดวัตถุประสงค์ที่กำหนดไว้ได้อย่างเหมาะสม ข้อทดสอบแต่ละแบบมีข้อดีและข้อเสียแตกต่างกัน จึงเป็นเรื่องที่ผู้สร้างแบบทดสอบควรจะได้ศึกษาเพิ่มเติมด้วย (Weir, 1990)

ขั้นที่ 4 เตรียมข้อทดสอบที่เหมาะสม (Preparing Relevant Test Items)

การเตรียมข้อทดสอบที่เหมาะสมมีขั้นตอนที่สำคัญ 3 ขั้นตอนคือ

ก. สร้างข้อทดสอบชนิดที่เหมาะสมกับวัตถุประสงค์ของการทดสอบ

เช่น หากต้องการข้อทดสอบที่สามารถทำให้เกิดความตรงและความเที่ยงสูงรวมทั้งง่ายต่อการตรวจ เนื่องจากใช้กับผู้สอบจำนวนมาก และคิดว่าเนื้อหาที่ต้องการทดสอบเหมาะสมกับอนุกรมวิธาน (Taxonomy) ของ Bloom ข้อทดสอบก็อาจเป็นแบบเลือกตอบหลายตัวเลือก ที่มุ่งทดสอบสิ่งที่ต้องการวัดดังนี้

1. ความรู้ความจำ

1.1 ความรู้เฉพาะอย่าง

ขั้นนี้เป็นขั้นตอนของความรู้เกี่ยวกับความหมายของคำศัพท์ นิยามของคำศัพท์ในศาสตร์และข้อเท็จจริงในศาสตร์ จึงควรทดสอบเกี่ยวกับนิยามศัพท์ ความหมายของศัพท์ สูตร กฎ หลักเกณฑ์ ความจริง ความสำคัญ ทฤษฎี คุณสมบัติเฉพาะ วัตถุประสงค์ สาเหตุและผลที่เกิดขึ้น รวมทั้งประโยชน์และโทษของเนื้อหาบางอย่างที่เรียน

1.2 ความรู้ในวิธีดำเนินการ

ขั้นนี้เป็นขั้นความรู้เกี่ยวกับระเบียบแบบแผน ลำดับขั้นของข้อความและแนวโน้มการจัดประเภทของความรู้ เกณฑ์ และวิธีดำเนินการที่เป็นขั้นตอนต่าง ๆ ลำดับ ขั้นตอนของข้อความ เหตุการณ์หรือปฏิบัติการ การจัดประเภทหรือชนิด ตัวอย่างของสิ่งที่เรียนเกณฑ์ในการจัดจำพวกสิ่งต่าง ๆ และวิธีการ (คือวิธี เทคนิค และกระบวนการปฏิบัติ) ในการดำเนินการเกี่ยวกับสิ่งที่เรียนไปแล้ว

1.3 ความรู้ในเรื่องหลักเกณฑ์ทั่วไปและความเป็นนามธรรม

ขั้นนี้เป็นขั้นความรู้เกี่ยวกับหลักการสำคัญ ข้อสรุปทั่วไป ทฤษฎีโครงสร้าง ความคิด และแนวคิดที่สำคัญของศาสตร์ซึ่งเป็นแก่นแท้ของความรู้ที่เรียนนั้น

2. ความเข้าใจ

ในขั้นนี้เป็นความรู้ 3 ระดับ คือ 1) การแปลความ ซึ่งเป็นการแปลความหมายของคำหรือข้อความจากระดับหนึ่งไปยังอีกระดับหนึ่ง เช่น การแปลความหมายของภาพ สูตร กราฟ ตาราง หรือวัตถุสิ่งของ และการเปรียบเทียบหรือเปรียบเปรย 2) การตีความเป็นการสรุปความจากการแปลความหมาย เช่น การตีความของข้อเท็จจริง หรือการเอาเรื่องราวเดิมมาคิดในแง่ใหม่ และ 3) การขยายความ เป็นการสรุปความจากการแปลความและการตีความในลักษณะของการอนุมาน เช่น $p = 0.10$ หมายความว่าอย่างไร (= ข้อทดสอบข้อหนึ่งมีผู้ตอบถูกเพียงร้อยละ 10 ของผู้เข้าสอบทั้งหมดในห้อง)

1. ข้อทดสอบยากเกินไป (แปลความ)
2. ข้อทดสอบนี้ไม่เหมาะสมกับเด็กในห้องนี้ (ตีความ)
3. ดังนั้นเด็กห้องนี้เสียเปรียบเด็กห้องอื่น (ขยายความ)

การขยายความนี้อาจทดสอบได้โดยการถามให้ผู้สอบขยายความด้านจินตนาการ การพยากรณ์ เหตุการณ์ การสมมติ และการอนุมาน เช่น ให้คาดคะเนหรือสรุปความจากรายละเอียดต่าง ๆ ของเรื่องที่เป็นอยู่ เป็นต้น

3. การนำไปใช้หรือการประยุกต์ คือการนำเอาสิ่งที่รู้แล้วไปใช้ในสถานการณ์ใหม่ที่คล้ายคลึงกัน ซึ่งเป็นการแก้ปัญหาเฉพาะ โดยอาศัยหลักวิชาที่เรียนรู้อาปรยุกต์เอง

4. การวิเคราะห์ เป็นความรู้ความสามารถในการแยกแยะสิ่งต่าง ๆ ออกเป็นส่วนย่อย เพื่อค้นหาความจริงบางอย่างที่ต้องการ ซึ่งการทดสอบอาจจะถามผู้สอนให้ทำการวิเคราะห์ความสำคัญ โดยนัย (implication) ความสัมพันธ์ของหลักการ โครงสร้าง และวิเคราะห์หลักเกณฑ์ในการจัดประเภทของสิ่งต่าง ๆ เป็นต้น

5. การสังเคราะห์ เป็นความรู้ความสามารถที่นำสิ่งต่าง ๆ มารวมกันเข้าอย่างเป็นระบบที่สัมพันธ์กันในลักษณะของความคิดสร้างสรรค์ (creativity) ซึ่งการทดสอบอาจทำได้โดยการถามเกี่ยวกับการสังเคราะห์ข้อความ การวางแผน การสรุปรวบยอดผลการศึกษาและการสังเคราะห์ความสัมพันธ์ของสิ่งต่าง ๆ เป็นต้น

6. การประเมินค่า เป็นความสามารถในการวินิจฉัยตีค่า ตีราคา และตัดสินใจอย่างมีหลักเกณฑ์ โดยอาศัยข้อความรู้ต่าง ๆ ที่เรียนรู้มาเป็นเกณฑ์ในการตัดสินใจ หรืออาศัยสิ่งที่สังคมยึดถือกันทั่วไปเป็นเกณฑ์ของความถูกต้อง

สำหรับคำที่แสดงการกระทำ (action words) ที่สามารถนำไปใช้ในการกำหนดพฤติกรรมเป้าหมายที่ต้องการทดสอบในแต่ละขั้นมีหลายคำ เช่น

พฤติกรรมด้านความรู้ เช่น บอก ชี้ เลือก นิยาม ให้ชื่อ จับคู่ ฯลฯ

พฤติกรรมด้านความเข้าใจ เช่น แปล อธิบาย ขยายความ ประมาณ อ้างอิง ยกตัวอย่าง ทำนาย สรุป ฯลฯ

พฤติกรรมด้านการนำไปใช้ เช่น คำนวณ เปลี่ยน สาริต ค้นหา คัดแปลง แก้ปัญหา จัดกระทำ ปฏิบัติ ผลิต ฯลฯ

พฤติกรรมด้านการวิเคราะห์ เช่น จัดแบ่ง หาความสัมพันธ์ หาความแตกต่าง จำแนก แยกวิเคราะห์ ชี้ความสำคัญ ฯลฯ

พฤติกรรมด้านการสังเคราะห์ เช่น จัดลำดับ จัดหมวดหมู่ สร้าง ประกอบแต่ง จัดระเบียบ ออกแบบ ย่อความ เรียบเรียง วางแผนการ ฯลฯ

พฤติกรรมด้านการประเมินค่า เช่น เปรียบเทียบ วิเคราะห์ ตัดสิน ประเมิน สรุปความ ฯลฯ

ข้อที่ควรระวังก็คือ ไม่ควรใช้คำซึ่งไม่แสดงพฤติกรรมเด่นชัดในข้อคำถามของข้อทดสอบ เพื่อกำหนดพฤติกรรมเป้าหมายที่ต้องการทดสอบ เช่น คำว่า ชอบ รู้สึก เข้าใจ ซาบซึ้ง ฯลฯ

ข. ปรับปรุงความตรงเชิงประจักษ์ (face validity) ของข้อทดสอบ

ความตรงเชิงประจักษ์ของข้อทดสอบเป็นเรื่องของความเหมาะสมของข้อทดสอบเท่าที่ปรากฏแก่สายตาว่าเป็นไปตามที่ควรจะเป็นหรือไม่ (Davies, 1979: 60) หรือหมายถึงหลักฐานต่าง ๆ เชิงประจักษ์ใด ๆ ที่ทำให้ผู้ใช้แบบทดสอบหรือข้อทดสอบมีความมั่นใจว่าแบบทดสอบนี้ใช้วัดสิ่งที่มุ่งทดสอบได้ ซึ่งนักทดสอบบางท่านเรียกว่าความตรงเชิงประจักษ์หรือความตรงที่ปรากฏ (empirical validity: Brown, 2000: 7)

แต่เนื่องจากชนิดของข้อทดสอบมีหลายประเภท ในที่นี้จะขอกล่าวถึงแต่เพียงข้อทดสอบแบบเลือกตอบเท่านั้นเพราะเป็นแบบที่นิยมกันแพร่หลาย ซึ่งในกรณีนี้การทำให้ข้อทดสอบมีความตรงเชิงประจักษ์สูงอาจทำได้หลายวิธี เช่น

1. สร้างตัวเลือกของข้อทดสอบแต่ละข้อให้มี 4 ตัวเลือก ไม่จำเป็นต้องมี 5 ตัว เลือกเพราะจะได้ผลไม่คุ้ม (Ebel, 1972: 205)
2. จัดรูปแบบ (visual format organization) ให้เหมาะสม เช่น เรียงตัวเลือกจากสั้นที่สุดไปหายาวที่สุด และหากเป็นจำนวนเลขก็ควรเรียงจากจำนวนที่น้อยที่สุดไปหามากที่สุดด้วย
3. เขียนคำสั่งให้ชัดเจนและครอบคลุม

ค. สร้างข้อทดสอบแบบเลือกตอบให้เป็นแบบทดสอบที่ดี

มีข้อเสนอแนะดังนี้ (Roid and Haladyna, 1982: 49-61; Gronlund, 1976: 158-209)

1. เขียนข้อทดสอบด้วยภาษาที่ถูกต้อง ชัดเจน และเข้าใจง่าย ไม่ใช่ข้อความหรือภาษาที่เข้าใจยากหรือตีความได้หลายนัย
2. เขียนข้อทดสอบให้อยู่ในรูปแบบเดียวกัน เช่น เป็นแบบคำถามหรือให้เติมข้อความให้ครบถ้วนก็ควรมีลักษณะอย่างเดียวกันตลอด
3. ไม่ควรเขียนข้อทดสอบเป็นแบบเลือกตอบซับซ้อน (multi-multiple-choice items) คือแบบทดสอบที่มีคำตอบที่ถูกที่สุดจากการรวมผสมกันระหว่างตัวเลือกต่าง ๆ เช่น 1) ข้อ ก. และ ข. 2) ข้อ ก. และ ค. 3) ข้อ ก. ข. และ ค. เป็นต้น
4. เขียนข้อทดสอบให้มีความยาวกะทัดรัดและกระชับความ ไม่เยิ่นเย้อโดยไม่จำเป็น
5. ตัวเลือก (choice) ของข้อทดสอบควรมีความยาวกะทัดรัดที่สุด และมีความยาวเท่า ๆ กัน (พยายามอย่าให้ตัวเลือกที่ถูกต้องยาวมากหรือสั้นมากกว่าตัวเลือกที่ผิด)
6. อย่าเขียนข้อทดสอบหรือตัวเลือกที่เป็นการชี้แนะ (clue) คำตอบระหว่างข้อหรือภายในข้อ
7. อย่าเขียนตัวเลือกให้มีความซ้ำซ้อนกัน แต่ควรเขียนให้แต่ละตัวเลือกมีความเป็นอิสระต่อกัน
8. อย่าใช้คำที่บ่งบอกความสมบูรณ์ที่สุด (absolute) ของเหตุการณ์หรือคำตอบ เช่น ใช้คำว่า ดีที่สุด ทุกคน ไม่เคยเลย เป็นต้น
9. กำบางคำไม่ควรนำมาใช้เป็นตัวเลือก เช่น ถูกทุกข้อ ผิดทุกข้อ และไม่ทราบ เป็นต้น
10. เมื่อต้องการเขียนข้อความที่เป็นปฏิเสธควรขีดเส้นใต้หรือพิมพ์ด้วยตัวเอน หรือตัวใหญ่ เพื่อให้เป็นที่สังเกตเห็นได้เด่นชัด
11. พยายามกระจายคำตอบที่ถูกของข้อทดสอบไว้ในตัวเลือกต่าง ๆ กันให้มีอัตราส่วนใกล้เคียงกัน ไม่ควรให้ข้อที่ถูกอยู่ที่ข้อใดข้อหนึ่งมากกว่าข้ออื่น ๆ เช่น คำตอบถูกไม่ควรจะเป็นข้อ ก. มากกว่าข้ออื่น ๆ เป็นต้น

12. เขียนโครงสร้างทางภาษาของตัวเลือกให้เหมือนกันทุกข้อหรืออย่างน้อยให้ใกล้เคียงกัน
13. ในแต่ละข้อให้มีตัวเลือกที่ถูกต้องเพียงข้อเดียวเท่านั้น (พยายามอย่าให้มีคำตอบถูกต้องมากกว่า 1 ข้อหรือไม่มีคำตอบที่ถูกต้องเลย)
14. พยายามเขียนตัวเลือกทุก ๆ ตัวเลือกให้มีลักษณะประหนึ่งว่าจะเป็นคำตอบที่ถูกต้องได้ทั้งนั้น
15. เขียนตัวเลือกทุก ๆ ตัวด้วยภาษาที่สอดคล้องสัมพันธ์กับข้อทดสอบ
16. ตัวเลือกไม่ควรมีความซ้ำกับข้อความในคำถามซึ่งเป็นข้อทดสอบ
17. ตัวเลือกที่เป็นตัวลวงควรเป็นตัวเลือกที่ทำให้ผู้สอบที่มีความสามารถต่ำเลือกตอบ มิใช่ทำให้ผู้สอบที่มีความสามารถสูงเลือกตอบ
18. ไม่ควรมีตัวเลือกที่ถูกต้องหรือไม่ถูกต้องที่ผู้สอบสามารถใช้ความรู้พื้นฐานทั่ว ๆ ไปตอบได้ว่าถูกหรือผิด และควรจะออกแบบกระดาษคำตอบให้ง่ายต่อการตรวจด้วย

ขั้นที่ 5 รวบรวมข้อทดสอบให้เป็นแบบทดสอบ (assembling the test)

ได้แก่การนำเอาข้อทดสอบแต่ละข้อรวบรวมจัดทำให้เป็นแบบทดสอบ ซึ่งผู้สร้างแบบทดสอบควรจะได้คำนึงถึงเรื่องการทำให้แบบทดสอบมีความตรงเชิงประจักษ์ดังที่ได้กล่าวมาแล้ว

ขั้นที่ 6 นำแบบทดสอบไปใช้ (administering the test)

นำแบบทดสอบสมิทธิภาพทั่วไปที่ได้จัดทำแล้วไปทดลองใช้จริงกับกลุ่มตัวอย่างผู้สอบที่มีจำนวนมากเพียงพอ (เช่น ประมาณ 300 คน) และเป็นตัวแทนที่ดีของประชากร เช่น มีความสามารถหลากหลายปะปนกันที่จะนำแบบทดสอบนี้ไปใช้จริงต่อไป

ขั้นที่ 7 ประเมินผลแบบทดสอบ (appraising the test)

ขั้นนี้มีความสำคัญมากที่สุดขั้นหนึ่งในการพัฒนาแบบทดสอบ ซึ่งอาจทำได้โดยการวิเคราะห์ข้อทดสอบรายข้อทางสถิติแล้วปรับปรุงข้อทดสอบรวมทั้งฉบับ (สุพัฒน์ สุขมลสันต์ 2538)

ขั้นที่ 8 นำผลการทดสอบไปใช้ (using the results)

ขั้นนี้เป็นขั้นสุดท้ายของการพัฒนาแบบทดสอบ ซึ่งได้แก่การตีความคะแนนที่ได้จากการทดสอบว่าหมายความว่าอย่างไร เพียงพอที่ผู้สอบจะนำความรู้ที่มีอยู่ไปใช้ให้เกิดประโยชน์ที่มุ่งหวังได้หรือไม่ และแน่ใจในการผลการตัดสินใจได้อย่างไร เป็นต้น

ดังนั้น จะเห็นได้ว่าการสร้างและการพัฒนาแบบทดสอบสมิทธิภาพทั่วไปนั้นมามีวิธีการและขั้นตอนหลายอย่างและแบบทดสอบที่จัดสร้างตามขั้นตอนต่าง ๆ ดังกล่าวแล้วปกติจะได้แบบทดสอบเพียงชุดเดียว

แต่หากผู้สร้างประสงค์จะสร้างแบบทดสอบหลายชุดที่มีลักษณะคู่ขนานกันทางเนื้อหา (equi-parallel forms) แล้วก็อาจทำได้ดังนี้

1. กำหนดวัตถุประสงค์ที่แน่ชัดเกี่ยวกับเนื้อหาที่ต้องการทดสอบ
2. ระบุผลลัพธ์ของข้อทดสอบ
3. สร้างข้อทดสอบและตัวเลือก ตามเนื้อหาและรูปแบบที่กำหนดไว้ตามขั้นตอนและวิธีการต่าง ๆ ดังได้กล่าวมาแล้วตั้งแต่ต้น

ตัวอย่าง

1. วัตถุประสงค์เชิงพฤติกรรม: บอกความหมายของคำศัพท์ที่เกี่ยวกับส่วนต่าง ๆ ของเรือใบได้
 2. ผลลัพธ์ของการทดสอบ: เลือกความหมายของคำศัพท์ที่เกี่ยวกับส่วนต่าง ๆ ของเรือใบได้ถูกต้อง
 3. รูปแบบของข้อทดสอบ: แบบเลือกตอบ
1. A large strong rope usually made of wire used to support a mast is a _____ .
 - a. jib
 - b. tuft
 - c. stay
 - d. gaff
 2. A sail outside a jib on an extension of the bowsprit is a _____ .
 - a. foresail
 - b. mainmast
 - c. stay sail
 - d. flying jib

นอกจากข้อทดสอบ 2 ข้อดังกล่าวแล้ว ผู้สร้างข้อทดสอบก็สามารถสร้างข้อทดสอบในลักษณะเดียวกันได้อีกเป็นจำนวนมากตามผลลัพธ์ของการทดสอบที่ได้กำหนดไว้ ทั้งนี้เพื่อวัตถุประสงค์เชิงพฤติกรรมที่ได้กำหนดไว้ เช่น อาจสร้างข้อทดสอบให้ผู้สอบบอกความหมายของส่วนอื่น ๆ ของเรือใบ เช่น jib, bowsprit, mainmast, foresail และ tuft เป็นต้น เทคนิควิธีการสร้างข้อทดสอบ (item form technique) นอกจากวิธีนี้แล้วยังมีวิธีอื่น ๆ อีก (Roid and Haladyna, 1982) เช่น

- Prose Transformation Technique
- Mapping-Sentence Technique
- Test Items for Concepts Technique
- LOGIQT (Logical Operations for Generating Intended Questions Technique)
- Instructional Quality Inventory Technique

ส่วนที่ 3 ความรู้เบื้องต้นเกี่ยวกับรูปแบบราช (Rasch Model)

1. แนวความคิดที่สำคัญเบื้องต้น (Lord, 1980; Willmott, 1974; Baker, 2001)

ปัจจุบันนี้ แนวคิดทางการทดสอบและประเมินผลมีความก้าวหน้ามากกว่าแต่ก่อน สาเหตุหนึ่งเกิดจากความก้าวหน้าทางเทคโนโลยีทางการศึกษาสมัยใหม่คือเครื่องคอมพิวเตอร์และความก้าวหน้าทางสถิติทางการศึกษา ทำให้แนวความคิดทางการทดสอบและประเมินผลก้าวหน้ามากขึ้นจนเกิดมีทฤษฎีทางการทดสอบใหม่ขึ้น เรียกว่า ทฤษฎีการตอบสนองต่อข้อทดสอบ (Item Response Theory) หรือที่รู้จักกันดีทั่วไปขณะนี้ว่าทฤษฎี IRT ทฤษฎีดังกล่าวนี้ มีหลายรูปแบบ เช่น

1. รูปแบบ 1 พารามิเตอร์ (One-parameter Model) หรือที่นิยมเรียกกันทั่วไปว่า Rasch Model เพื่อเป็นเกียรติแก่ Georg Rasch นักคณิตศาสตร์ชาวเดนมาร์กที่คิดค้นขึ้น รูปแบบนี้ใช้ค่าความยาก (difficulty index: β) ในสูตรคำนวณหาคุณภาพของข้อทดสอบ
2. รูปแบบ 2 พารามิเตอร์ (Two-parameter Model) อาศัยค่าพารามิเตอร์ 2 ค่าในการหาคุณภาพของข้อทดสอบ คือค่าความยาก และค่าอำนาจจำแนก (discrimination index: α)
3. รูปแบบ 3 พารามิเตอร์ (Three-parameter Model) อาศัยค่าพารามิเตอร์ 3 ค่าในการหาคุณภาพของข้อทดสอบ คือค่าความยาก ค่าอำนาจจำแนก และค่าการเดาถูก (guessing index: c)
4. รูปแบบ 4 พารามิเตอร์ (Four-parameter Model) อาศัยค่าพารามิเตอร์ 4 ค่าในการหาคุณภาพของข้อทดสอบ คือค่าความยาก ค่าอำนาจจำแนก ค่าการเดาถูก และค่าความคลาดเคลื่อน (error index: ξ)

ในการกำหนดเกณฑ์ความสามารถของผู้สอบแบบทดสอบนั้น ผู้สร้างแบบทดสอบอาจอาศัยรูปแบบต่าง ๆ ของทฤษฎี IRT ดังกล่าวแล้วได้ ยกเว้นรูปแบบ 4 พารามิเตอร์ เพราะเป็นรูปแบบใหม่ล่าสุดที่เป็นเพียงแนวคิดและสูตรในการคำนวณหาค่าพารามิเตอร์ต่าง ๆ ยังมีปัญหาอยู่ ปกติแล้ว การคำนวณหาคุณภาพของแบบทดสอบหรือข้อทดสอบโดยอาศัยรูปแบบ 2 พารามิเตอร์ และ 3 พารามิเตอร์ เสียค่าใช้จ่ายมากเพราะว่ามักต้องอาศัยกลุ่มตัวอย่างประชากรขนาดใหญ่ (คือมากกว่า 1,000 คน) และใช้เวลาในการคำนวณหาค่าพารามิเตอร์นานกว่าจะได้ค่าลู่เข้า (convergent values) ที่คงที่ ดังนั้น นักทดสอบทั่วไปหรือครู-อาจารย์ทั่วไป จึงนิยมใช้รูปแบบ 1 พารามิเตอร์มากกว่าเพราะว่าสามารถใช้ได้กับกลุ่มตัวอย่างขนาดเล็กได้ดีกว่า เช่น ประมาณ 100 คน ก็ใช้ได้และเสียเวลารวมทั้งค่าใช้จ่ายในการดำเนินงานเพื่อให้ได้ค่าพารามิเตอร์ที่คงที่น้อยกว่ามาก

ในงานวิจัยนี้ ผู้วิจัยสนใจที่จะนำรูปแบบราชหรือรูปแบบ 1 พารามิเตอร์ มาใช้เพื่อวิเคราะห์หาค่าความยาก (β) ของข้อทดสอบรายข้อ ซึ่งเป็นค่าที่ไม่เปลี่ยนแปลงไปตามลักษณะของผู้สอบ (person-free) ดังนั้น ค่าความยากนี้จะประ โยชน์ในการประกอบการพิจารณาจัดเรียงข้อทดสอบจากข้อที่ง่ายที่สุดไปหาข้อที่ยากที่สุดเมื่อนำแบบทดสอบไปใช้จริงต่อไป

ในทางทฤษฎีการทดสอบแล้ว การที่จะสร้างแบบทดสอบใด ๆ นั้น สิ่งที่คุณสร้างแบบทดสอบควรต้องคำนึงถึงมากที่สุด คือ ความตรงตามสภาวะสันนิษฐาน (construct validity) ซึ่งได้แก่แนวคิดเบื้องหลังในเชิงทฤษฎีว่าแบบทดสอบนั้นจะวัดอะไร มากน้อยเพียงไร และทำไมจึงต้องวัดสิ่งต่าง ๆ เหล่านั้น เช่น ในกรณีของการทดสอบความสามารถในการอ่านเข้าใจความภาษาอังกฤษ (reading comprehension) นั้น จะต้องวัดความสามารถด้านใดบ้างจึงจะทำให้ผลการทดสอบสะท้อนออกมาว่าผู้สอบมีความรู้ความสามารถในการอ่านเข้าใจความ ผู้สร้างแบบทดสอบอาจจะต้องการวัดความสามารถ ต่าง ๆ เช่น การอ้างอิงโดยตรง (direct reference) การอนุมานความหมาย (inference) การสมมุติ (supposition) และการประเมินผล (evaluation) เป็นต้น (Davis and Widdowson, 1974: 168-176) และในบางระดับผู้สร้างแบบทดสอบอาจต้องการวัดความสามารถอื่น เช่น การถอดความ (paraphrasing) การทดสอบคำศัพท์ในและนอกบริบท (vocabulary in and out of context) และการสรุปเนื้อเรื่อง (conclusion) เป็นต้น และเมื่อกำหนดขอบเขตของความตรงตามสภาวะสันนิษฐานได้แล้ว จึงกำหนดความตรงเชิงเนื้อหา (content validity) ตามประเด็น ปัญหา ก็คือ เมื่อนำแบบทดสอบไปใช้แล้ว ผู้ทำการทดสอบจะทราบได้อย่างไรว่า ข้อทดสอบแต่ละข้อสามารถทำหน้าที่ในการวัดสิ่งที่มุ่งจะทำการวัดได้ และข้อที่ต้องการจะวัดสิ่งเดียวกันสามารถวัดสิ่งเดียวกันได้หรือไม่ การที่จะหาคำตอบดังกล่าวนี้ได้ นักทดสอบอาจทำได้หลายอย่าง เช่น (Warm, 1978)

1. วิเคราะห์องค์ประกอบ (Factor Analysis) ของผลการทดสอบเพื่อตรวจสอบว่าข้อทดสอบที่มุ่งทำการวัดความสามารถเดียวกัน สามารถจับกลุ่มรวมกันได้หรือไม่ วิธีนี้เป็นวิธีที่ง่ายในเชิงการคำนวณ เพราะในปัจจุบันนี้อาจจะสามารถใช้เครื่องคอมพิวเตอร์ช่วยได้ แต่ในด้านการดำเนินการมีความจำกัด กล่าวคือ วิธีการวิเคราะห์องค์ประกอบมีหลายวิธี แต่ละวิธีมีแนวคิดเชิงทฤษฎีแตกต่างกันทำให้ผลการวิเคราะห์ที่แตกต่างกัน นอกจากนี้ต้องอาศัยกลุ่มตัวอย่างผู้สอบจำนวนมาก คือ ประมาณ 10 เท่าของจำนวนข้อทดสอบ ซึ่งในระยะการพัฒนาเครื่องมือทดสอบมักเป็นปัญหาในการดำเนินการ
2. วิธีหาความสัมพันธ์ระหว่างข้อของผลการทดสอบ (correlation matrix) เพื่อตรวจสอบว่าข้อทดสอบที่วัดสิ่งเดียวกันมีความสัมพันธ์กันสูงหรือไม่ วิธีนี้ง่ายในเชิงการคำนวณ แต่ไม่สามารถตอบได้ว่าข้อทดสอบแต่ละข้อมีความยากง่ายแตกต่างกันหรือไม่ ทำให้ไม่สามารถทราบได้ว่า ผลของคะแนนรวมแต่ละข้อมีความหมายว่าอย่างไร
3. วิเคราะห์ผลการทดสอบตามทฤษฎีการตอบสนองต่อข้อทดสอบ (Item Response Theory) ซึ่งมีหลายวิธี

สิ่งที่วิธีการวิเคราะห์ต่าง ๆ ดังกล่าวแล้วข้างต้นนี้มุ่งแสวงหาก็คือ 1) ลำดับที่ของความยากของข้อทดสอบ (item difficulty order) ทั้งฉบับ เมื่อคำนึงถึงค่าพารามิเตอร์ต่าง ๆ ที่เกี่ยวข้องและ 2) การรวมกลุ่มกันขององค์ความรู้ที่มุ่งทำการทดสอบ ปัจจัยทั้ง 2 ประการนี้ เป็นสิ่งสำคัญมากในการกำหนดเกณฑ์

ความสามารถของผู้สอบแต่ละคน และการวิเคราะห์ข้อทดสอบด้วยรูปแบบราชจะได้คำตอบทั้งสองประการ ดังกล่าวนี้ด้วย

2. รูปแบบราช (Rasch Model)

1. ประวัติความเป็นมาโดยสังเขป (Willmott, 1980; Wright and Panchapakesan, 1969)

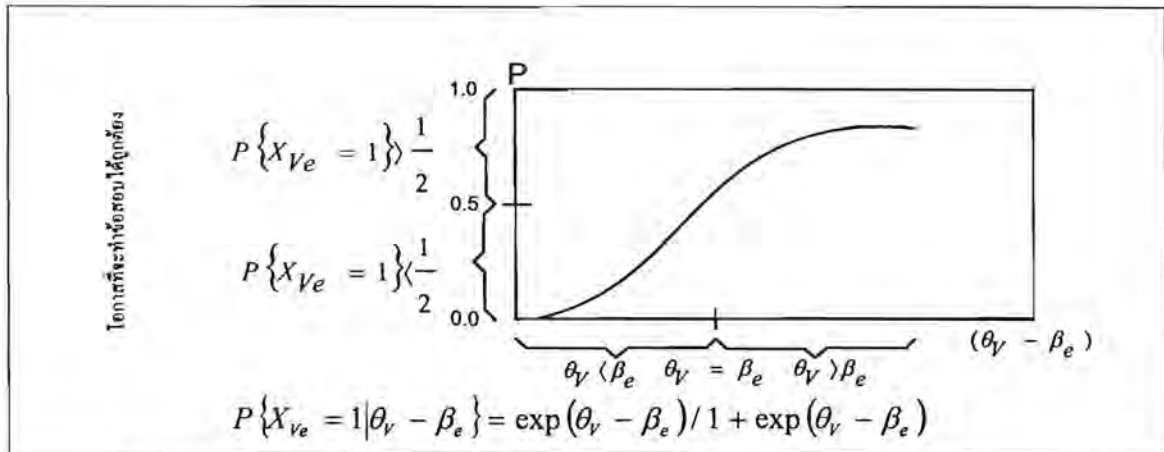
ในปี 1960 Georg Rasch นักคณิตศาสตร์ชาวเดนมาร์กได้เสนอแนวคิดเกี่ยวกับการทดสอบตามรูปแบบ 1 พารามิเตอร์ขึ้น และได้ตีพิมพ์บทความในทำนองเดียวกันอีกในปี 1961 และ 1966 ต่อมา Benjamin Wright ได้เผยแพร่แนวคิดนี้ให้เป็นที่แพร่หลายทั่วไปตั้งแต่ปี 1967 เป็นต้นมา และเรียกชื่อแนวคิดนี้ว่า Rasch Model แนวคิดที่สำคัญของ Rasch Model ก็คือแนวคิดที่เชื่อว่าความเป็นปรนัยของการวัดผล (objectivity of measurement) ที่ไม่สามารถจะหาได้จากทฤษฎีการทดสอบแบบเดิม (Classical Test Theory) ทั่ว ๆ ไป สามารถจะหาได้จากการทดสอบตามรูปแบบราชคือ

1. ความเป็นอิสระจากกลุ่มตัวอย่าง (sample-free test calibration) กล่าวคือ การคำนวณหาค่าพารามิเตอร์ของข้อทดสอบเป็นอิสระจากกลุ่มตัวอย่าง ค่าต่าง ๆ ของข้อทดสอบ เช่น ค่าความยาก (b) จะไม่แปรเปลี่ยนไปตามลักษณะกลุ่มตัวอย่าง และ
2. ความเป็นอิสระจากข้อทดสอบ (item-free person measurement) กล่าวคือ การคำนวณหาค่าพารามิเตอร์ของบุคคลเป็นอิสระจากข้อทดสอบ เช่น ความสามารถของผู้สอบ (θ) จะไม่แปรเปลี่ยนไปตามลักษณะของข้อทดสอบ ความสามารถของผู้สอบจะคงที่ไม่ว่าจะวัดด้วยข้อทดสอบใด ข้อทดสอบหนึ่งที่ตนเคยทำเลขแล้ว ไม่ว่าข้อทดสอบนั้นจะไปปรากฏอยู่ที่ส่วนใดของแบบทดสอบ

แนวความคิดดังกล่าวนี้ นับว่ามีประโยชน์มากในการทดสอบและวัดผล โดยเฉพาะการทดสอบเกี่ยวกับความสามารถทางสมอง (mental measurement) ในปี 1969 Wright และ Panchapakesan ได้ทำการทดสอบแนวคิดดังกล่าวแล้ว และสนับสนุนว่าเป็นไปได้จริงและในปีต่อมา มาได้มีนักทดสอบได้ทำการทดสอบแนวคิดดังกล่าวอีกหลายครั้ง และยืนยันว่า คุณสมบัติดังกล่าวนี้เป็นไปได้จริงแม้แต่กลุ่มตัวอย่างจะไม่ได้จากการสุ่มมาก็ตาม และมีความแกร่งมาก (robustness) คือสามารถทดสอบเพื่อคำนวณหาความสามารถของผู้สอบ (θ) และความยากของข้อทดสอบ (β) ได้ทั้ง ๆ ที่กลุ่มตัวอย่างบุคคลและข้อทดสอบแตกต่างจากข้อตกลงเบื้องต้นของการใช้รูปแบบนี้ก็ตาม เช่น ขนาดของกลุ่มตัวอย่างมีเพียง 100 คนก็ได้ และไม่จำเป็นต้องมากจากการสุ่มก็ได้ (Wright, 1979: 219)

2. แนวคิดของรูปแบบราช

ตามแนวคิดของรูปแบบราชนั้น โอกาสที่คนจะทำข้อทดสอบได้หรือไม่ขึ้นอยู่กับระดับความสามารถของตนเอง (ability parameter; θ) และระดับความยากของข้อทดสอบ (difficulty index: b) เช่น ถ้า $\beta = .5$ และ $\theta = .5$ โอกาสที่คนผู้นั้นจะสามารถทำข้อทดสอบข้อนั้นได้ถูกต้องประมาณ 50% ถ้าหากความสามารถของผู้สอบ (θ) น้อยกว่าความสามารถของข้อทดสอบ (ความยากหรือ β) แล้ว โอกาสที่จะตอบข้อทดสอบนั้นได้ถูกต้องก็ย่อมจะน้อยกว่า 50% และในทำนองเดียวกัน ถ้าหากว่าค่า θ มากกว่าค่า β แล้ว โอกาสที่คนผู้นั้นจะตอบข้อทดสอบได้ถูกต้องก็มีมากกว่า 50% (ดูภาพประกอบที่ 3)



ภาพประกอบที่ 2: ความสัมพันธ์ระหว่าง θ_v และ β_e (หรือ b_e) ของตัวแปรและโค้งลักษณะเฉพาะของข้อทดสอบ (item characteristic curve: ICC) ของรูปแบบราช

ดังนั้น ตามแนวคิดของ Rasch Model นั้น ค่าพารามิเตอร์ที่เกี่ยวข้องมีเพียง 2 ค่าเท่านั้น ไม่มีอำนาจจำแนก หรือโอกาสของการเดาถูกเข้ามาเกี่ยวข้องด้วย เพราะเป็นข้อตกลงเบื้องต้นว่าข้อทดสอบที่นำมาใช้กับรูปแบบราชนั้น จะต้องมีความจำแนกเท่า ๆ กันและมีลักษณะที่จะทำให้เกิดการเดาได้น้อยที่สุด แต่ในทางปฏิบัติลักษณะทั้งสองนี้อาจไม่เป็นจริง แต่ว่าความแกร่ง (robustness) ของรูปแบบนี้สามารถจะแก้ไขปัญหาได้

3. พัฒนาการของสูตรของรูปแบบราช

ดังได้กล่าวมาแล้ว แนวคิดของรูปแบบราช เกี่ยวข้องกับความสัมพันธ์ของความสามารถของผู้สอบ (θ) และความยากของข้อทดสอบ (β) เท่านั้น ความสัมพันธ์นี้เป็นความสัมพันธ์กับเชิงคณิตศาสตร์ กล่าวคือ

โอกาสที่บุคคล V ที่มีระดับความสามารถ θ (หรือ θ_v) จะทำข้อทดสอบ e ที่มีระดับความยากง่าย β (หรือ β_e) จะตอบข้อทดสอบได้ถูกมากน้อยเพียงใดขึ้นอยู่กับระดับความแตกต่างของ $(\theta_v - \beta_e)$ นั่นคือ

โอกาสของความสำเร็จ (odd of success) = $(\theta_v - \beta_e)$ นั้น มีค่าระหว่าง $\pm\infty$ แต่ค่าโอกาสของความสำเร็จมีได้ระหว่าง 1 กับ 0 เท่านั้น เพื่อให้ค่าของ $(\theta_v - \beta_e)$ เป็นค่าที่มีหน่วยเล็กลง และคงที่เหมาะสมแก่การนำมาใช้ จึงใช้ค่า exponent ของ $(\theta_v - \beta_e)$ แทน ซึ่งจะมีค่าอยู่ระหว่าง 0 กับ $+\infty$ เช่น

$$\begin{aligned}\text{odd of success} &= e^{(\theta_v - \beta_e)} \\ &= \text{EXP}(\theta_v - \beta_e) \dots\dots\dots 1\end{aligned}$$

หน่วยของการวัดเรียกว่า log-odds scale หรือ logits และ ค่า $e = 2.71818$

เพื่อทำให้ค่า $\text{EXP}(\theta_v - \beta_e)$ มีค่าเป็นมาตรวัดอันตรภาค (interval scale) ระหว่าง 0 และ 1 สมการที่ 1 อาจเขียนได้ดังนี้

$$\text{odd of success} = \text{EXP}(\theta_v - \beta_e) / [1 + \text{EXP}(\theta_v - \beta_e)] \dots\dots\dots 2$$

ดังนั้น สูตรที่แสดงว่าโอกาสที่บุคคล V ที่มีความสามารถ θ จะทำข้อทดสอบ e ที่มีความยาก b ได้ถูกต้อง (ได้คะแนน = 1 หรือ X_{vi}) คือ

$$P \{X_{vi}=1/\theta_v, \beta_e\} = \text{EXP}(\theta_v - \beta_e) / [1 + \text{EXP}(\theta_v - \beta_e)] \dots\dots\dots 3$$

นี่คือสูตรของรูปแบบราช ตัวอย่างเช่น

ถ้า $(\theta_v - \beta_e) = 2$ log - odds หน่วย โอกาสที่บุคคลจะทำข้อทดสอบข้อที่ e ได้ถูกต้องคือ

$$\begin{aligned}P \{X_{vi}=1/\theta_v, \beta_e\} &= e^2 / [1+e^2] \\ &= 2.71828^2 / [1+2.71828^2] \\ &= .88 \text{ หรือ } 88.1\%\end{aligned}$$

ความสัมพันธ์ระหว่างโอกาสที่จะตอบข้อทดสอบได้ถูกต้องกับระดับความยากของข้อทดสอบ มีการกระจายเป็นรูปคล้ายตัวเอส (ogive distribution) ต่อมาเพื่อให้การกระจายของโค้งดังกล่าวเป็นโค้งรูปตัวเอสปกติ (normal ogive) จึงนิยมใช้ค่า 1.7 เข้ามาเกี่ยวข้องในการคำนวณด้วยเพื่อปรับโค้งของความสัมพันธ์นี้ ปกติค่าที่นำมาปรับนี้ทำให้ความแตกต่างของโค้งทั้ง 2 ลักษณะแตกต่างกันน้อยกว่า 1%

4. ข้อตกลงเบื้องต้นในการใช้รูปแบบราช

ข้อตกลงเบื้องต้นในการใช้รูปแบบราชมีดังนี้ (Hambleton and Cook, 1977: 78; Tinsley and Davis, 1975: 326)

1. ข้อทดสอบแต่ละข้อจะต้องวัดคุณลักษณะ (trait) เดียวกัน (unidimensional latent space) กล่าวคือ ข้อทดสอบจะต้องเป็นเอกพันธ์ (homogeneous items) และวัดเอกมิติ (unidimensionality) ในการที่จะตัดสินว่าข้อทดสอบมีลักษณะดังกล่าวหรือไม่ อาจทำได้ทำได้หลายวิธี เช่น (Warm, 1978: 101)
 - ก. วิธี Eigenvalue Test โดยการวิเคราะห์องค์ประกอบ (Factor Analysis) ของข้อทดสอบทั้งหมด หากปรากฏว่า eigenvalue ที่ 1 มีค่าสูงและแตกต่างจาก eigenvalue ที่ 2 มาก และ eigenvalue ที่ 2 แตกต่างจากค่า eigenvalue อื่น ๆ ไม่มาก แสดงว่าแบบทดสอบนั้นวัดเอกมิติ
 - ข. วิธี Baseline Test โดยการหาความสัมพันธ์ระหว่างค่า Factor loading ของ factor ที่ 1 กับค่า Biserial correlation ของข้อทดสอบรายข้อกับคะแนนรวม หากว่าค่าสัมประสิทธิ์สหสัมพันธ์ > 0.80 แสดงว่าแบบทดสอบนั้นวัดเอกมิติ

อนึ่ง ถ้าข้อทดสอบเป็นวิวิธพันธุ์ (heterogeneous items) จะต้องมีการวิเคราะห์ดังกล่าวเพื่อจัดกลุ่มเอกพันธุ์ก่อน

2. ข้อทดสอบจะต้องมีความเป็นอิสระจากตำแหน่ง (local independence) ใน 2 ลักษณะคือ
 - ก. มีความเป็นอิสระทางสถิติ (Statistical independence) กล่าวคือข้อทดสอบแต่ละข้อเป็นอิสระไม่เกี่ยวข้องกัน แต่ละข้อวัดความสามารถที่ไม่ซ้ำกันเลย ดังนั้น คำตอบของแต่ละข้อเป็นอิสระต่อกัน แต่รวมกันแล้วจะวัดคุณลักษณะเดียวกันเท่านั้น
 - ข. มีความเป็นอิสระเสรี (uncorrelated independence) กล่าวคือ ข้อทดสอบแต่ละข้อจะปรากฏอยู่ในตำแหน่งใดของข้อทดสอบก็ได้ จะไม่มีผลต่อการตอบของผู้ตอบข้อทดสอบก็ได้ จะไม่มีผลต่อการตอบของผู้ตอบข้อทดสอบ
 - ค. คะแนนมี 2 ลักษณะเท่านั้น คือเป็น 1 หรือ 0
 - ง. การตอบข้อทดสอบได้ถูกต้องหรือไม่ ขึ้นอยู่กับความสัมพันธ์ของความสามารถของผู้สอบกับความยากของข้อทดสอบเท่านั้น

5. ความแกร่ง (robustness) ของรูปแบบราช

ในการใช้รูปแบบราชเพื่อการทดสอบนั้น ได้มีการทดลองใช้การทดสอบในรูปแบบที่ต่าง ๆ กันเพื่อหาความแกร่งของรูปแบบนี้ ปรากฏว่ารูปแบบนี้ได้รับความนิยมมาก เพราะว่ามี ความแกร่งมากดังนี้คือ

1. ขนาดของกลุ่มตัวอย่างที่จะใช้กับรูปแบบราช อาจจะมีขนาดเล็กประมาณ 100 คน เท่านั้นก็ได้ (Wright, 1977: 219)

2. การกระจายของความสามารถของกลุ่มตัวอย่างหรือความยากของข้อทดสอบไม่จำเป็นต้องเป็น โค้งปกติ เพียงแต่มีแนวโน้มว่าจะเป็น โค้งปกติเท่านั้นก็ได้ และตัวอย่างไม่จำเป็นที่จะได้มาเพราะการสุ่มก็ได้ (Wright and Stone, 1979: 20)
3. ข้อทดสอบ ไม่จำเป็นต้องมีอำนาจจำแนกเท่ากันก็ได้ และไม่จำเป็นต้องคำนึงถึงค่าการเดาก็ได้ (Wright and Panchapakesan, 1969: 25)
4. ข้อทดสอบที่ให้คะแนนหลากหลายมากกว่า 0, 1 (multiple-point) ก็ใช้กับรูปแบบราชได้ (Willmott, 1980: 195)

6. วิเคราะห์ผลการทดสอบรายข้อตามรูปแบบราช

ผลการทดสอบรายข้อของแบบทดสอบปรนัยแบบเลือกตอบ (multiple-choice test) สามารถนำมาวิเคราะห์รายข้อตามรูปแบบราชได้ โดยใช้โปรแกรมคอมพิวเตอร์ เช่น ใช้โปรแกรม BICAL/PC (Binomial Calibration for Personal Computer) โปรแกรม Logist ของ ETS (Educational Testing Service) โปรแกรม Quest ของ ACER (Australian Centre for Education Research) และโปรแกรม ASCAL ของบริษัท MicroCAT แห่งอเมริกา และโปรแกรม Rasch (สุพัฒน์ สุขมลสันต์, 2542) เป็นต้น โปรแกรมเหล่านี้จะให้ค่าสถิติที่สำคัญหลายอย่างแต่ที่สำคัญมาก คือ

1. ค่าความยากของข้อทดสอบรายข้อ (β) ที่เรียงจากข้อที่ง่ายที่สุดไปหาข้อที่ยากที่สุด
2. แผนภาพ (chart) แสดงการกระจายของค่าความยากของข้อทดสอบรายข้อและค่าความถี่ของการกระจายซึ่งมักเรียกกันว่าแผนที่ตัวแปร (map of variables) และ
3. แผนภาพแสดงการตอบข้อทดสอบถูกและผิดของผู้ตอบแต่ละคนในการตอบข้อทดสอบแต่ละข้อ ซึ่งมักเรียกว่าแผนที่ผู้ตอบข้อทดสอบ (kid map)

ผลการวิเคราะห์ข้อทดสอบข้อที่ 1 มีประโยชน์ในการจัดเรียงข้อทดสอบ ส่วนข้อที่ 2 มีประโยชน์ในการกำหนดเกณฑ์ระดับความสามารถของผู้สอบข้อทดสอบ และข้อที่ 3 มีประโยชน์ในการตรวจสอบเกณฑ์ที่ผู้สร้างแบบทดสอบกำหนดขึ้น

ส่วนที่ 4 ความสำคัญและลักษณะเฉพาะของโปรแกรมที่จะพัฒนาขึ้นใช้

1. โปรแกรมวิเคราะห์ข้อทดสอบแบบประเพณีนิยม (Classical Test Model) (สุพัฒน์ สุขมลสันต์, 2542)

ก. ความสำคัญของการวิเคราะห์ข้อทดสอบแบบประเพณีนิยม

การวิเคราะห์ข้อทดสอบแบบประเพณีนิยมใช้เพื่อวิเคราะห์หาคุณภาพของข้อทดสอบแบบเลือกตอบแต่ละข้อและแบบทดสอบทั้งฉบับ เหมาะสำหรับการแบบทดสอบอิงกลุ่ม (Norm-referenced Test)

และอาจใช้ได้กับแบบทดสอบแบบอิงเกณฑ์ (Criterion-referenced Test) หรืออิงปริเขต (Domain-referenced Test) ได้ด้วย การวิเคราะห์ข้อทดสอบจะได้ค่าสถิติที่จำเป็นดังนี้

1) ค่าความยากของข้อทดสอบ (Item-Difficulty) เพื่อต้องการทราบว่าข้อทดสอบแต่ละข้อมีความยากมากน้อยเพียงไร เหมาะแก่การนำไปใช้ต่อไปหรือไม่

2) ค่าอำนาจจำแนกของข้อทดสอบ (Item Discrimination Power) เพื่อต้องการทราบว่าข้อทดสอบแต่ละข้อนั้นมีอำนาจจำแนกผู้ที่เรียนเก่ง และเรียนอ่อนออกจากกันได้มากน้อยเพียงใด

3) ประสิทธิภาพของตัวเลือก (Effectiveness of Distractors) นอกจากตัวเลือกที่ถูกต้องแล้ว ยังต้องการทราบว่าตัวเลือก หรือ ตัวลวง ของข้อทดสอบแต่ละข้อนั้นมีลักษณะที่ดีหรือไม่ ถ้าไม่ดี จะได้หาทางปรับปรุงให้ดีขึ้นต่อไป

4) ความเที่ยงของแบบทดสอบ (Reliability of the Test) เพื่อต้องการทราบว่าข้อทดสอบแต่ละข้อมีความเที่ยงมากน้อยเพียงใด

5) ค่าสถิติทั่วไปของแบบทดสอบ (General Test Statistics) เราต้องการทราบว่าความเที่ยงของแบบทดสอบมีมากน้อยเพียงใด ค่าความยากของแบบทดสอบเป็นอย่างไร และค่าอำนาจจำแนกโดยเฉลี่ยของแบบทดสอบสูงหรือต่ำเพียงใด เพื่อเป็นแนวทางในการตัดสินใจได้ว่าแบบทดสอบของเราเป็นอย่างไรบ้าง เช่น มีความเชื่อถือได้มากน้อยเพียงไร เพื่อเป็นแนวทางในการปรับปรุงแบบทดสอบให้ดีขึ้นในครั้งต่อไป

ข. ลักษณะเฉพาะของโปรแกรมวิเคราะห์ข้อทดสอบแบบประเพณีนิยมที่จะพัฒนาขึ้นใช้

1. ใช้การวิเคราะห์ข้อทดสอบด้วย 2 วิธี คือ วิธีเทคนิค 27% และวิธีเทคนิค 50% ซึ่งมีรายละเอียดในการใช้ดังนี้

1) เทคนิค 27%

วิธีนี้นิยมใช้เมื่อผู้สอบแบบทดสอบทั้งหมดมีจำนวนมาก เช่น มากกว่า 1,000 คน หรือการกระจายของคะแนนสอบเป็นโค้งปกติ (Normal Curve) หรือมีแนวโน้มว่าเป็นโค้งปกติ ซึ่งวิธีนี้เชื่อว่ากลุ่มตัวอย่างขนาด 27% จากผู้สอบที่ได้คะแนนสูงหรือผู้สอบที่ได้คะแนนต่ำสามารถเป็นตัวแทนประชากรของผู้สอบทั้งหมดได้

2) เทคนิค 50%

วิธีนี้นิยมใช้กับการวิเคราะห์ข้อทดสอบที่มีผู้สอบจำนวนน้อย เช่น ประมาณ 30-100 คน ซึ่งปกติแล้วคะแนนสอบมักกระจายไม่เป็นโค้งปกติ จึงต้องใช้ผลการทดสอบทั้งหมดมาทำการวิเคราะห์

2. มีค่าสถิติที่สำคัญ 2 ประเภทดังต่อไปนี้

2.1 ค่าสถิติที่เกี่ยวกับลักษณะของข้อทดสอบแต่ละข้อ มีหลายอย่าง คือ

1) ค่าความยากของแบบทดสอบ

ค่าความยากของแบบทดสอบ หมายถึงอัตราส่วนของจำนวนผู้ตอบข้อทดสอบนั้นสามารถตอบได้ถูกต้องอย่างน้อยเท่าใด แบ่งได้เป็น 4 แบบ คือ

1.1) ค่าความยากของข้อทดสอบสำหรับกลุ่มเก่ง ซึ่งการหาค่าความยากโดยวิธีนี้อาศัยเทคนิค 27% ของผู้ที่สอบได้คะแนนสูง โดยใช้สูตรดังนี้

$$p = \frac{N_u}{N_1}$$

เมื่อ p = ระดับความยากของตัวเลือกแต่ละข้อสำหรับผู้สอบกลุ่มเก่ง

N_u = จำนวนคนในกลุ่มเก่งจำนวน 27% ที่เลือกตัวเลือก i

N_1 = จำนวนคนทั้งหมดในกลุ่มเก่ง 27%

ค่านี้ให้ผลลัพธ์เรียกว่า ดัชนีความยากของข้อทดสอบสำหรับผู้สอบกลุ่มเก่ง (Difficulty Index for Upper Group)

1.2) ค่าความยากของข้อทดสอบสำหรับกลุ่มอ่อน ซึ่งการหาค่าความยากโดยวิธีนี้อาศัยเทคนิค 27% ของผู้ที่สอบได้คะแนนต่ำ โดยใช้สูตรดังนี้

$$p = \frac{N_l}{N_2}$$

เมื่อ p = ระดับความยากของตัวเลือกแต่ละข้อสำหรับผู้สอบกลุ่มอ่อน

N_l = จำนวนคนในกลุ่มอ่อนจำนวน 27% ที่เลือกตัวเลือก i

N_2 = จำนวนคนทั้งหมดในกลุ่มอ่อน 27%

ค่านี้ให้ผลลัพธ์เรียกว่า ดัชนีความยากของข้อทดสอบสำหรับผู้สอบกลุ่มอ่อน (Difficulty Index for Lower Group)

1.3) ค่าความยากของข้อทดสอบสำหรับผู้สอบทั้งหมด การหาค่าความยากวิธีนี้ใช้เทคนิค 50% คืออาศัยจำนวนผู้สอบทั้งหมดที่เข้าสอบ โดยใช้สูตรดังนี้

$$p = \frac{N_u}{N_t}$$

เมื่อ p = ค่าความยากของตัวเลือกแต่ละข้อสำหรับผู้สอบทั้งหมด

N_u = จำนวนผู้สอบทั้งหมดที่เลือกตอบตัวเลือก i

[ในที่นี้ N_i คือจำนวนผู้ตอบตัวเลือก i เกิดจากจำนวนคนในกลุ่มเก่ง (N_g) + จำนวนคนในกลุ่มอ่อน (N_o)]

N_r = จำนวนผู้สอบทั้งหมด

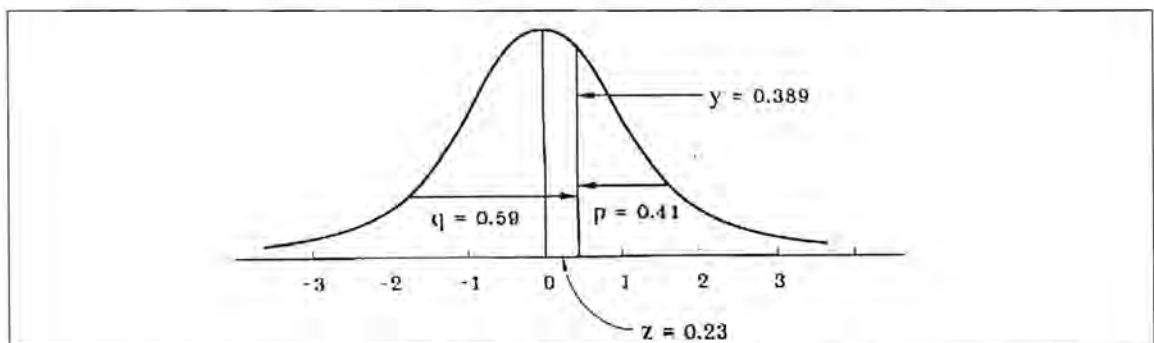
ค่านี้ให้ผลลัพธ์เรียกว่า ดัชนีความยากรวมของข้อทดสอบ (Total Difficulty Index)

ดัชนีความยาก (p) ของวิธีต่างๆดังกล่าวนี้มีค่าระหว่าง 0-1.0 ถ้าค่า p ยิ่งมาก ข้อทดสอบนั้นยิ่งง่าย แต่ในทางกลับกัน ถ้าค่า p ยิ่งน้อย ก็แสดงว่าข้อทดสอบนั้นยิ่งยาก ข้อทดสอบที่ดี ควรมีค่า p อยู่ระหว่าง 0.20-0.80

1.4) ค่าความยากแบบเคลตา เนื่องจากค่า p เป็นค่าที่ขัดกับความรู้สึกที่ว่าข้อทดสอบที่มีค่า p มาก จะง่าย แต่ข้อทดสอบที่มีค่าน้อยจะยาก และค่า p มีมาตราส่วน (scale) ไม่เท่ากัน เพราะไม่ใช่ค่าที่เกิดจากการคำนวณพื้นที่ภายใต้โค้งปกติ บางครั้งจำนวนผู้ตอบข้อทดสอบแต่ละข้อไม่เท่ากันเพราะ บางข้อมีเว้นไว้ไม่ตอบ (omitted) หรือทำแบบทดสอบไม่เสร็จ (unreached) ค่าเคลตาเป็นค่าที่คำนวณจากคะแนนมาตรฐานของจำนวนผู้ตอบข้อนั้นถูกหรือผิดจากสมการต่อไปนี้

$$\Delta = 13 + 4z$$

เมื่อ z = ค่ามาตรฐานของโค้งปกติ ณ จุด p (อัตราส่วนร้อยละของผู้ที่ตอบข้อทดสอบข้อนั้น ได้ถูกต้อง) เช่น เมื่อ $p = 0.41$, $z = 0.23$ และ จะได้ $\Delta = 13.92$ ดังแสดงในรูปที่ 2.1



ภาพประกอบที่ 3: แสดงที่มาของค่า z และ y

ค่าความยากของแบบทดสอบ (p) มีการกำหนดค่าความยากไว้ดังนี้

ตารางที่ 8: แสดงรายละเอียดของค่าความยาก

ค่าความยาก	ระดับ
0.4-0.6	ดีมาก
0.31-0.39 , 0.61-0.69	ดี
0.20-0.30 , 0.70-0.80	พอใช้ได้
0.80 >	ง่ายเกินไป
< 0.19	ยากเกินไป

จากสมการสามารถสรุปได้ว่า ข้อทดสอบที่มีค่าเคลตา(Δ)มากจะมีความยากมากกว่าข้อทดสอบที่มีค่าเคลตาน้อย ค่านี้สามารถหาค่าเฉลี่ย หรือนำมาเปรียบเทียบกันได้

2) ค่าอำนาจจำแนกของข้อทดสอบ

ค่าอำนาจจำแนกของข้อทดสอบ หรือค่าความตรงของข้อทดสอบ (Item Validity) คือ ดัชนีที่ชี้ว่า ข้อทดสอบแต่ละข้อมีความสามารถในการจำแนกผู้สอบที่เก่งออกจากผู้สอบที่ไม่เก่ง ได้มากน้อยเพียงใด หรือเป็นดัชนีที่แสดงถึงความสัมพันธ์ระหว่างข้อทดสอบแต่ละข้อกับคะแนนรวม หรือสิ่งที่มุ่งทดสอบว่ามีมากน้อยเพียงใด ค่าอำนาจจำแนกของข้อทดสอบมีอยู่ 3 วิธี คือ

2.1) ดัชนีการจำแนก (Discrimination Index :Disc.Index) คำนวณได้จากสมการต่อไปนี้

$$d = \frac{N_u - N_l}{N}$$

เมื่อ d = ค่าอำนาจจำแนก

N_u = จำนวนคนในกลุ่มเก่งจำนวน 27% ที่เลือกตัวเลือก i

N_l = จำนวนคนในกลุ่มอ่อนจำนวน 27% ที่เลือกตัวเลือก i

N = จำนวนคนทั้งหมดในกลุ่ม 27%

ค่าที่คำนวณได้จะอยู่ระหว่าง 0 ถึง 1.0 ถ้าจำนวนผู้สอบในกลุ่มเก่งและกลุ่มอ่อนเลือกตอบตัวเลือกเดียวกันจำนวนเท่ากัน ค่า $d = 0$ แสดงว่าตัวเลือกดังกล่าวไม่สามารถจำแนกคนเก่งกับคนไม่เก่งออกจากกันได้ แต่หากว่ามีผู้สอบจากกลุ่มอ่อนเลือกตัวเลือกนั้นมากกว่าผู้สอบจากกลุ่มเก่ง ค่า d จะติดลบ ดังนั้นตัวเลือกที่ไม่ใช่คำตอบที่ถูกต้องของข้อทดสอบแต่ละข้อควรมีค่า d ติดลบ ส่วนตัวเลือกที่ถูกต้องไม่ควรจะมีค่า d ติดลบ

ค่าอำนาจจำแนกของแบบทดสอบ (d) มีความหมายดังนี้

ตารางที่ 9: แสดงรายละเอียดของค่าอำนาจจำแนก

ค่าอำนาจจำแนก	ระดับ
0.6 >	ดีมาก
0.35-0.59	ดี
0.20-0.34	พอใช้ได้
< 0.20	ควรปรับปรุง

2.2) ค่าสหสัมพันธ์แบบไบซีเรียล (Biserial Correlation: r_{bis})

การหาค่าอำนาจการจำแนกของข้อทดสอบวิธีนี้ ตั้งอยู่บนสมมติฐานที่ว่าคะแนนของผู้ที่ตอบข้อทดสอบถูกและผิดแต่ละข้อนั้นมีการกระจายเป็น โค้งปกติ ค่าสหสัมพันธ์ระหว่างผลรวมของคะแนนทั้ง 2 ชุดนี้กับคะแนนรวมก็กระจายเป็น โค้งปกติเช่นเดียวกัน แสดงถึงค่าอำนาจจำแนกของข้อทดสอบแต่ละข้อรวมทั้งตัวเลือกตอบอื่นๆอีกด้วย ค่า r_{bis} มีความหมายเช่นเดียวกับค่าสหสัมพันธ์แบบเพียร์สัน (Pearson Product-Moment Correlation: r_{xy}) และเป็นค่าโดยประมาณของ r_{xy} ด้วย ตามปกติค่า r_{bis} จะอยู่ระหว่าง 0 ถึง 1 เช่นเดียวกับค่า r_{xy} แต่ถ้าหากผลรวมของคะแนนที่ใช้เป็นเกณฑ์ (คะแนนรวม) ไม่เป็น โค้งปกติ หรือว่ามียอด 2 ยอด (bimodal curve) หรือเบ้มาก ค่า r_{bis} อาจจะมีค่ามากกว่า 1.00 ก็ได้ ซึ่งค่า r_{bis} สามารถคำนวณได้จากสมการ

$$r_{bis} = \frac{\bar{x}_r - \bar{x}_w}{S.D._r} \cdot \frac{p(1-p)}{y}$$

เมื่อ \bar{x}_r = ค่าเฉลี่ยของคะแนนผู้สอบที่ตอบข้อทดสอบนั้นถูก

\bar{x}_w = ค่าเฉลี่ยของคะแนนผู้สอบที่ตอบข้อทดสอบนั้นผิด

$S.D._r$ = ค่า S.D. ของคะแนนรวมของผู้สอบทั้งหมด

p = อัตราส่วนของผู้ที่ตอบข้อทดสอบนั้นถูก

y = ความสูงของแกน y ใน โค้งปกติ ณ จุดที่ค่า p

จากสมการข้างต้นจะสังเกตได้ว่า ตัวเลือกที่ผิดจะพิจารณารวมกัน แล้วค่อยแยกคำนวณทีละตัวเลือกในภายหลัง ซึ่งค่าอาจจะไม่ถูกต้องเลยทีเดียว

2.3) ค่าสหสัมพันธ์แบบพอยต์ไบซีเรียล (Point-Biserial Correlation: r_{pb})

การคำนวณคล้ายกับการคำนวณค่า r_{bis} แต่ค่า r_{pb} นั้นใช้เมื่อการกระจายของคะแนนรวมหรือคะแนนตัวเลือกที่ถูกหรือผิดไม่เป็นโค้งปกติ โดยใช้สมการดังนี้

$$r_{pb} = \frac{\bar{x}_r - \bar{x}_w}{S.D._r} \cdot \sqrt{p(1-p)}$$

ปกติค่า r_{pb} จะเท่ากับค่าสหสัมพันธ์แบบเพียร์สัน เมื่อข้อที่ถูกมีค่า = 1 และข้อที่ผิดมีค่า = 0 และตามปกติค่า r_{pb} จะมีค่าน้อยกว่า r_{bis} เสมอ

นอกจากนี้ เพื่อทดสอบว่าค่าอำนาจจำแนกแต่ละชนิดของแบบทดสอบมีนัยสำคัญหรือไม่ จึงทำการทดสอบค่าสัมประสิทธิ์สหสัมพันธ์ของแต่ละชนิดด้วย t-test โดยอาศัยสมการดังต่อไปนี้

$$t = \frac{r_{xz}}{\sqrt{(1-r_{xz}^2)/(n-2)}} ; df = (n-2)$$

เมื่อ r_{xz} = ค่าความเที่ยงแบบ r_{pb} หรือ r_{bis}

n =

จำนวนผู้สอบทั้งหมดที่ตอบข้อทดสอบนั้น

ค่า t ที่คำนวณได้จะมีนัยสำคัญหรือไม่ สามารถเปรียบเทียบค่าที่คำนวณได้กับค่าจากตารางค่าวิกฤตของสัมประสิทธิ์สหสัมพันธ์ (Critical value of the correlation coefficient) จากตำราสถิติทั่วไปได้ เช่น เมื่อ $df > 100$ ถ้า $t > 1.96$ ค่า t ซึ่งเป็นค่าอำนาจจำแนกนี้จะมีนัยสำคัญที่ $p = 0.05$ เป็นต้น

3) ค่าเกณฑ์เฉลี่ย (Mean criterion)

นอกจากค่าสหสัมพันธ์ทั้ง 3 ชนิด ดังที่กล่าวมาแล้ว ค่าเฉลี่ยของเกณฑ์ (คะแนนรวม) สามารถใช้เป็นดัชนีบ่งบอกถึงประสิทธิภาพของตัวเลือกแต่ละตัวของข้อทดสอบแต่ละข้อ เพราะเป็นค่าที่แสดงว่าผู้ที่เลือกตอบตัวเลือกแต่ละตัวมีคะแนนทั้งหมดในการทดสอบข้อทดสอบชุดนั้นเท่าใด ถ้าค่าเฉลี่ยของเกณฑ์ของตัวเลือกที่ถูกต้องน้อยกว่าค่าเฉลี่ยของเกณฑ์ของตัวลวงแล้วแสดงว่า ข้อทดสอบข้อนั้นยังไม่ดีพอ เพราะมีนักเรียนจำนวนมาก หรือนักเรียนเก่งเลือกตอบตัวลวงเป็นคำตอบที่ถูกต้อง ข้อทดสอบข้อนั้นจึงควรปรับปรุงซึ่งเกณฑ์ค่าเฉลี่ย มี 2 อย่างคือ

3.1) คะแนนเกณฑ์เฉลี่ย (Mean criterion score) ตัวเลือกที่ดีของข้อทดสอบแต่ละข้อ ควรมีความหมาย (Score) ประมาณ $\bar{X} + ISE_{\bar{X}}$ (Standard error of mean) ถ้าคะแนนมากกว่าค่าดังกล่าวแสดงว่าตัวเลือกนั้นง่าย แต่ถ้าคะแนนนั้นน้อยกว่าค่านี้แสดงว่าตัวเลือกนั้นยาก เป็นต้น

3.2) คะแนนมาตรฐานที่ (T-Score) ตัวเลือกแต่ละตัวของข้อทดสอบแต่ละข้อจะมีคะแนนมาตรฐานที่เชิงเส้น (linear T-score) กำกับไว้ด้วย เพื่อชี้ให้เห็นว่าผู้ที่เลือกตัวเลือกแต่ละตัวนั้นได้คะแนนรวมโดยเฉลี่ยแล้วเป็นค่ามาตรฐานที่เชิงเส้นเท่าใด ตัวเลือกที่ดีที่สุดคือ ตัวเลือกที่ถูกต่อนั้นผู้เลือกควรได้คะแนนมาตรฐานที่เชิงเส้นมากที่สุด จึงจะถือได้ว่าเป็นตัวเลือกที่ดี

2.2 ค่าทางสถิติที่สำคัญสำหรับแบบทดสอบทั้งหมด

ค่าสถิติที่เกี่ยวข้องกับแบบทดสอบทั้งหมด มี 3 ลักษณะ คือ

1) ค่าสถิติบรรยายพื้นฐาน

1.1) ค่ามัธยฐาน (Median)

ค่ามัธยฐาน ได้แก่คะแนนที่ร้อยละ 50 ของผู้สอบทั้งหมดได้คะแนนมากกว่าคะแนนดังกล่าว และมีผู้สอบอีกร้อยละ 50 ที่ได้คะแนนน้อยกว่าคะแนนดังกล่าว ในกรณีที่ผู้สอบมีจำนวนคี่ ค่ามัธยฐานจะได้แก่คะแนนจุดกึ่งกลางระหว่างคะแนนที่เรียงมาจกมากไปหาน้อย หรือจากน้อยไปหามาก อาจกล่าวได้ว่าค่ามัธยฐานได้แก่ค่าลำดับชั้นเปอร์เซ็นต์ไทล์ที่ 50 (50th Percentile Rank: P₅₀) หรือค่าลำดับ ควอร์ไทล์ที่ 2 (2nd Quartile Rank: Q₂) ซึ่งสามารถคำนวณได้จากสมการ

$$Md = l + \frac{\frac{n}{2} - cf}{f} - i$$

โดยที่ l = ซีดจำกัดล่าง (Lower limit) ของชั้นคะแนนที่มีค่ามัธยฐานอยู่

n = จำนวนผู้เข้าสอบทั้งหมด

cf = ค่าความถี่สะสมของชั้นคะแนนก่อนคะแนนค่ามัธยฐาน

f = ค่าความถี่ของคะแนนในชั้นคะแนนที่ค่ามัธยฐานอยู่

i = ค่าอินตรัภาค (interval) ของชั้นคะแนน

1.2) ค่าฐานนิยม (Mode)

ค่าฐานนิยม ได้แก่คะแนนที่มีความถี่มากที่สุด หรือกล่าวอีกนัยหนึ่งก็คือคะแนนที่มีผู้สอบสอบได้จำนวนมากที่สุดของการทดสอบในแต่ละครั้ง ซึ่งสามารถคำนวณได้จากสมการ

$$\text{Mode} = 3 \text{ Median} - 2 \text{ Mean}$$

1.3) คะแนนต่ำสุด (Minimum Score)

ได้แก่ ค่าคะแนนต่ำที่สุดที่มีผู้สอบแต่ละครั้งสอบได้

1.4) คะแนนสูงสุด (Maximum Score)

ได้แก่ คะแนนสูงที่สุดที่มีผู้สอบแต่ละครั้งสอบได้

1.5) พิสัย (Range)

ได้แก่ค่าความแตกต่างระหว่างคะแนนสูงสุดและคะแนนต่ำสุด ตามสมการ

$$\text{Range} = \max - \min$$

1.6) ค่าเบี่ยงเบนควอร์ไทล์ (Quartile Derivation)

ค่าเบี่ยงเบนควอร์ไทล์ได้แก่ค่าเฉลี่ยความแตกต่างของคะแนนสอบระหว่างผู้ที่ได้คะแนนในตำแหน่งควอร์ไทล์ที่ 3 (Q_3) หรือเปอร์เซ็นต์ไทล์ที่ 75 (P_{75}) กับคะแนนสอบของผู้ที่ได้คะแนนในตำแหน่งควอร์ไทล์ที่ 1 (Q_1) หรือเปอร์เซ็นต์ไทล์ที่ 25 (P_{25}) หรือกล่าวอีกนัยหนึ่งก็คือค่าเฉลี่ยของความแตกต่างระหว่างคะแนนของผู้ที่สอบได้คะแนนมากกว่าผู้อื่นร้อยละ 75 กับคะแนนของผู้ที่สอบได้คะแนนมากกว่าผู้อื่นร้อยละ 25 เป็นเท่าใด ซึ่งสามารถคำนวณได้จากสมการ

$$QD = \frac{Q_3 - Q_1}{2}$$

1.7) ค่าความเบ้ (Skewness)

ความเบ้ คือลักษณะของการขาดความสมมาตร (symmetry) ของการกระจายของคะแนนเมื่อผู้เข้าสอบจำนวนมากว่าครึ่งหนึ่งได้คะแนนมากกว่าหรือน้อยกว่าคะแนนเฉลี่ย ถ้าผู้สอบจำนวนมากว่าครึ่งหนึ่งสอบได้คะแนนมากกว่าคะแนนเฉลี่ย การกระจายของคะแนนจะเบ้ไปทางลบ (negatively skewed curve) แต่ถ้าผู้สอบจำนวนมากว่าครึ่งหนึ่งสอบได้คะแนนน้อยกว่าคะแนนเฉลี่ย การกระจายของคะแนนจะเบ้ไปทางบวก เรียกว่าโค้งเบ้ทางบวก (positively skewed curve) ซึ่งสมการแสดงการคำนวณการหาความเบ้สามารถแสดงได้คือ

$$Sk = \frac{m_3}{m_2^3}$$

โดยที่ m_3 = โมเมนต์ที่ 3 (third moment) จะหาได้จากค่าของ

$$\frac{\sum (X - \bar{X})^3}{n}$$

m_2 = โมเมนต์ที่ 2 (second moment) จะหาได้จากค่าของ

$$\frac{\sum (X - \bar{X})^2}{n}$$

Sk มีค่าอยู่ระหว่าง ± 3.0 ถ้า $Sk = 0.00$ แสดงว่าการกระจายของคะแนนมี

ลักษณะสมมาตร

1.8) ค่าความโค้ง (Kurtosis)

ความโค้ง คือ ลักษณะความสูงของโค้งการกระจายของคะแนนว่ามีลักษณะแหลม (peaked) หรือแบนลาด (flat-topped) หรือว่าปกติ (normal) ซึ่งสามารถหาค่าความโค้งได้จากสมการนี้

$$Ku = \frac{m_4}{m_2^2}$$

โดยที่ m_4 = โมเมนต์ที่ 4 (fourth moment) หาได้จากสมการต่อไปนี้

$$\frac{\sum(X - \bar{X})^4}{n}$$

ค่าความโค้งมีความหมายดังนี้ คือ

- ก. ถ้าค่า $Ku > 3$ แสดงว่าการกระจายของคะแนนมีลักษณะเป็นโค้งแหลม (leptokurtic)
- ข. ถ้าค่า $Ku = 3$ แสดงว่าการกระจายของคะแนนเป็นโค้งปกติ (mesokurtic or normal curve)
- ค. ถ้าค่า $Ku < 3$ แสดงว่าการกระจายของคะแนนมีลักษณะเป็นโค้งแบนลาด (platykurtic)

2) ค่าสรุปทางสถิติทั่วไปของแบบทดสอบ

ได้แก่ ค่าเฉลี่ย ค่าต่ำสุด ค่ามัธยฐาน ค่าเบี่ยงเบนมาตรฐาน และค่าความแปรปรวนของค่าต่างๆ ดังต่อไปนี้คือ

- 2.1) คะแนนของแบบทดสอบ
- 2.2) ดัชนีค่าความยาก
- 2.3) เฉลี่ย
- 2.4) ดัชนีการจำแนก
- 2.5) ค่าสหสัมพันธ์แบบไบซีเรียล (biserial correlation)
- 2.6) ค่าสหสัมพันธ์แบบพอยต์ไบซีเรียล (point-biserial correlation)
- 2.7) ดัชนีความเที่ยงแบบต่างๆ

3) ค่าความเที่ยงของแบบทดสอบ (test reliability) (สุพรรณ สุกลมลันต์, 2542)

ค่าความเที่ยงของแบบทดสอบ เป็นแบบความสอดคล้องภายใน (internal consistency) ซึ่งสามารถคำนวณได้ 4 วิธี (พร้อมทั้งค่าความคลาดเคลื่อนมาตรฐานในการวัด (standard error of measurement : SE_m หรือ SEM) ดังนี้

1. Kuder-Richardson Formula 20 (KR_{20})

การคำนวณหาความเที่ยงแบบนี้ อาศัยอัตราส่วนของผู้ที่สอบได้และสอบตกในการตอบข้อทดสอบแต่ละข้อเป็นสำคัญ ซึ่งถ้าทำการคำนวณด้วยมือจะมีความยุ่งยากมากจึงไม่ค่อยมีผู้นิยมนำมาใช้ แต่ว่าเป็นค่าที่ถูกต้องมากที่สุดสำหรับแบบทดสอบเพื่อการจำแนก (power test) ที่ข้อทดสอบมีลักษณะเป็นวิวิธพันธ์ (heterogeneous) ที่มุ่งทดสอบหลายอย่างโดยการถามในหลายๆด้าน และมีจำนวนข้อมาก เช่น แบบทดสอบสมิทธิภาพ (proficiency test) เป็นต้น ค่าความเที่ยงแบบนี้สามารถคำนวณได้จากสมการ

$$KR_{20} = \left[\frac{K}{K-1} \right] \left[1 - \frac{\sum p_i \cdot q_i}{S.D.^2} \right]$$

เมื่อ K = จำนวนข้อทดสอบทั้งหมด

p_i = อัตราส่วนของผู้ที่ตอบถูกแต่ละข้อ

$$q_i = (1 - p_i)$$

$S.D.^2$ = ความแปรปรวนของคะแนนสอบของผู้สอบทั้งหมด

2. Kuder-Richardson formula 21 (KR_{21})

การคำนวณหาความเที่ยงวิธีนี้เป็นง่ายกว่าแบบ KR_{20} มากเพราะใช้เพียงค่า \bar{X} , $S.D$ และค่า k เท่านั้น ดังนั้นค่า KR_{21} จึงเป็นที่นิยมใช้กันทั่วไป แต่ค่า KR_{21} นี้มีความถูกต้องน้อยกว่าค่า KR_{20} และมีค่าน้อยกว่าค่า KR_{20} ด้วย ซึ่งค่า KR_{21} นี้เหมาะสมกับแบบทดสอบที่ข้อทดสอบแต่ละข้อมีค่าความยากเท่าๆกัน โดยที่ค่า KR_{21} นี้สามารถคำนวณได้จากสมการ

$$KR_{21} = \frac{k}{k-1} \left[1 - \frac{\bar{X}(k - \bar{X})}{k.S.D.^2} \right]$$

3. Cronbach alpha (α)

ในการคำนวณหาค่าความเที่ยงแบบนี้ ต้องอาศัยค่าความแปรปรวนในการตอบข้อทดสอบแต่ละข้อ ซึ่งเกิดจากผลคูณของอัตราส่วนในการตอบถูกและตอบผิดของผู้สอบที่นำมาวิเคราะห์เช่นเดียวกับการหาค่า KR_{20} และในกรณีที่การให้คะแนนสอบแต่ละข้อเป็น 1, 0 คือถ้าตอบถูกจะได้ 1 คะแนน แต่ถ้าตอบผิดจะได้ 0 คะแนน ซึ่งค่าแอลฟานี้จะเท่ากับค่า KR_{20} พอดี สำหรับการหาค่าความเที่ยงแบบนี้อาจจะใช้กับแบบทดสอบหรือแบบทดสอบถามที่ให้คะแนนมากกว่า 1 คะแนน (multiple point) ได้ด้วย กล่าวคือ ข้อที่ตอบถูกอาจจะมีคะแนนที่มากกว่า 1 คะแนน และมีหลายค่า เช่น 3, 2, 1 เป็นต้น แต่สำหรับข้อที่ตอบผิด หรือไม่ตอบจะมีค่าเป็น 0 โดยที่ค่าแอลฟาสามารถคำนวณได้จากสมการต่อไปนี้

$$\alpha = \frac{k}{k-1} \left[1 - \frac{\sum S.D._i^2}{S.D.^2} \right]$$

เมื่อ $S.D._i^2$ = ความแปรปรวนของคะแนนสอบแต่ละข้อ

$S.D.^2$ = ความแปรปรวนของคะแนนสอบของผู้สอบทั้งหมด

4. วิธีแบ่งครึ่งแบบทดสอบ (split-half method: r_u)

ค่าความเที่ยง r_u คือค่าสหสัมพันธ์ระหว่างข้อทดสอบข้อคู่และข้อคี่ของแบบทดสอบชุดเดียวกัน ค่าความเที่ยงชนิดนี้จะมากกว่าค่า KR_{20} และ KR_{21} หรือค่า Cronbach alpha และอาจกล่าวได้ว่าเป็นค่าความเที่ยงที่มีขนาดใหญ่มากที่สุด ค่า r_u นี้จึงมีความเหมาะสมกับแบบทดสอบเพื่อการจำแนก (power test) ที่มีข้อทดสอบเป็นแบบวิธีพื้นฐานเช่นเดียวกัน ค่า r_u สามารถคำนวณได้จากสมการ

$$r_u = \frac{2r_{xy}}{1 + r_{xy}}$$

เมื่อ r_{xy} = ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนข้อคู่และข้อคี่ ที่ได้มาจากการคำนวณแบบสหสัมพันธ์อย่างง่ายของเพียร์สัน

อนึ่ง จะสังเกตได้ว่าการหาค่าความเที่ยงแบบความสอดคล้องภายในทั้ง 4 วิธีดังกล่าวไม่เหมาะสมกับแบบทดสอบเพื่อการแข่งขัน (speed test) ที่ข้อทดสอบมีลักษณะเป็นเอกพันธ์ (homogeneous) ที่มุ่งทดสอบสิ่งเดียวกันแต่จะถามในหลายๆด้าน และมีจำนวนข้อมาก เช่น แบบทดสอบวินิจฉัย (diagnostic test) เพราะว่าการทดสอบแบบนี้ ปกติผู้สอบมักทำข้อทดสอบได้ไม่ครบทุกข้อ ทำให้เราไม่สามารถทราบจำนวนข้อทดสอบที่ผู้สอบได้ทำครบทั้งหมดทุกคนมีจำนวนเท่าใด ดังนั้นถ้าใช้การคำนวณด้วย 4 วิธีดังกล่าว

จะทำให้ได้ค่าความเที่ยงที่สูงมากเกินความเป็นจริง ในกรณีที่เป็นแบบทดสอบเพื่อการแข่งขันควรวัดค่าความเที่ยงโดยวิธีการทดสอบซ้ำ (test-retest method) หรือใช้วิธีการทดสอบแบบคู่ขนาน (equivalent-form method) จะได้ค่าที่มีความถูกต้องมากกว่า ส่วนการคำนวณหาค่าความคลาดเคลื่อนมาตรฐานในการวัด (standard error of measurement) ของทั้ง 4 วิธีนั้น สามารถคำนวณได้จากสมการนี้

$$SEM = S.D. \cdot \sqrt{1 - r_{xx}}$$

เมื่อ $r_{xx} = KR_{20}, KR_{21}, \alpha$ หรือ ค่า r_{rr}

2. โปรแกรมวิเคราะห์ข้อทดสอบแบบราช (Rasch Model) (สุพัตน์ สุกมลสันต์, 2542; ศิริชัย กาญจนวาสี, 2545)

ก. ความสำคัญของการวิเคราะห์ข้อทดสอบแบบราช

1. ใช้วิเคราะห์ข้อทดสอบปรนัยแบบเลือกตอบตามแนวคิดของทฤษฎีการตอบสนองต่อข้อทดสอบ (Item Response Theory) ซึ่งเชื่อว่าความเป็นปรนัยของการวัดผล (Objectivity of measurement) ที่ไม่สามารถจะหาได้จากแบบประเพณีนิยมทั่วไปคือ

1). ความเป็นอิสระจากกลุ่มตัวอย่าง (Sample-Independence) กล่าวคือ การคำนวณหาค่าพารามิเตอร์ของข้อทดสอบเป็นอิสระจากกลุ่มตัวอย่าง ค่าต่างๆ ของข้อทดสอบ เช่น ค่าความยากของข้อทดสอบ (β) จะไม่แปรเปลี่ยนไปตามกลุ่มลักษณะของกลุ่มตัวอย่าง

2). ความเป็นอิสระจากข้อทดสอบ (Item-Independence) กล่าวคือ การคำนวณหาค่าพารามิเตอร์ของบุคคลเป็นอิสระจากข้อทดสอบ เช่น ความสามารถของผู้สอบ (θ) จะไม่แปรเปลี่ยนไปตามลักษณะของการทดสอบ ความสามารถของผู้สอบจะคงที่ไม่ว่าจะวัดเมื่อใด เช่น ข้อทดสอบข้อหนึ่งๆ ที่เคยทดสอบแล้วจะมีค่าคงที่เมื่อวัดกับบุคคลเดิม ไม่ว่าข้อทดสอบนั้นจะไปปรากฏอยู่ส่วนใดของแบบทดสอบ

ตามแนวความคิดของการวิเคราะห์ข้อทดสอบแบบราชนั้นเชื่อว่า โอกาสที่ผู้สอบจะตอบข้อทดสอบได้ถูกต้องหรือไม่ ขึ้นอยู่กับปัจจัยที่สำคัญ 2 อย่างคือ ระดับความสามารถของตนเอง (ability parameter, θ) และค่าความยากของข้อทดสอบ (difficult parameter, β) เช่น ถ้า $\theta = 0.5$ และ $\beta = 0.5$ โอกาสที่คนผู้นั้นจะสามารถตอบข้อทดสอบนั้นได้ถูกต้องมี 50% ถ้าหากความสามารถของผู้สอบ (θ) น้อยกว่าความสามารถของข้อทดสอบ (ค่าความยากหรือ β) แล้วโอกาสที่คนผู้นั้นจะตอบข้อทดสอบนั้นได้ถูกต้องย่อมจะน้อยกว่า 50% และในทำนองเดียวกัน ถ้าหากว่าค่า θ มากกว่าค่า β แล้วโอกาสที่คนผู้นั้นจะตอบข้อทดสอบได้ถูกต้องก็มีมากกว่า 50%

ดังนั้น ตามแนวความคิดของการวิเคราะห์ข้อทดสอบแบบราช ค่าพารามิเตอร์ที่เกี่ยวข้องมีเพียง 2 ค่าเท่านั้น คือ ค่าความสามารถของผู้สอบ (θ) และค่าความยากของข้อทดสอบ (β) ไม่มีค่าอำนาจจำแนก หรือโอกาสของการเดาเข้ามาเกี่ยวข้องด้วย ความสัมพันธ์เหล่านี้เป็นความสัมพันธ์เชิงคณิตศาสตร์ กล่าวคือ โอกาสที่บุคคล v ที่มีระดับความสามารถ θ (หรือ θ_v) จะสามารถทำข้อทดสอบ e ที่มีค่าความยาก β หรือ β_e ได้ถูกต้องอย่างน้อยเพียงใดนั้นขึ้นอยู่กับระดับความแตกต่างของ $(\theta_v - \beta_e)$ นั่นคือ

$$\text{โอกาสของความสำเร็จ (Odd of success)} = (\theta_v - \beta_e)$$

แต่เนื่องจาก $(\theta_v - \beta_e)$ มีค่าอยู่ที่ระหว่าง $\pm \infty$ แต่ค่าโอกาสของความสำเร็จมีได้ระหว่าง 1 กับ 0 เท่านั้น เพื่อให้ค่าของ $(\theta_v - \beta_e)$ เป็นค่าที่มีหน่วยเล็กลงและคงที่เหมาะสมแก่การนำมาใช้ จึงใช้ค่าเลขชี้กำลัง (exponent) ของ $(\theta_v - \beta_e)$ แทน ซึ่งจะมีค่าอยู่ระหว่าง 0 กับ $\pm \infty$

$$\text{โอกาสของความสำเร็จ} = e^{(\theta_v - \beta_e)} = \exp(\theta_v - \beta_e)$$

หน่วยของการวัดเรียกว่า log-odds unit หรือ ลอจิตส์ (logits) เพื่อให้ค่า $\exp(\theta_v - \beta_e)$ มีค่าเป็นมาตราอันตรภาค (interval scale) ระหว่าง 0 และ 1 ดังนั้นสมการข้างต้นอาจเขียนใหม่ได้ดังนี้

$$\text{โอกาสความสำเร็จ} = \frac{\exp(\theta_v - \beta_e)}{1 + \exp(\theta_v - \beta_e)}$$

ดังนั้นสมการที่แสดงว่าโอกาสที่บุคคล v มีความสามารถ θ จะตอบข้อทดสอบข้อที่ e ที่มีค่าความยาก β ได้ถูกต้อง (ได้คะแนน = 1 หรือ X_v) คือ

$$P\{X_v = 1 | \theta_v, \beta_e\} = \frac{\exp(\theta_v - \beta_e)}{1 + \exp(\theta_v - \beta_e)}$$

2. ค่าสถิติที่คำนวณได้เป็นค่าที่ไม่เปลี่ยนแปลงไปตามลักษณะของกลุ่มตัวอย่าง ทำให้มีประโยชน์มากกว่าค่าที่ได้จากการวิเคราะห์ด้วยแบบประเพณีนิยม เช่น

- 1) ใช้เก็บไว้ในธนาคารข้อทดสอบ (item bank)
- 2) ใช้ในการทดสอบปรับเปลี่ยน (adaptive testing)
- 3) ใช้กำหนดระดับความสามารถ (ability band setting)
- 4) ใช้สร้างแบบทดสอบที่ให้คะแนนรายข้อต่างกัน (multiple-point test)
- 5) ใช้กำหนดเกณฑ์ของระดับสัมฤทธิ์ (mastery level) ของแบบทดสอบแบบอิงเกณฑ์
- 6) ใช้วินิจฉัยความสามารถของผู้ตอบ (diagnostic)
- 7) ใช้หาความเป็นอคติของข้อทดสอบ (item bias)

- 8) ใช้เปรียบเทียบคะแนนต่างชุด (equating scores)
- 9) ใช้ศึกษาปัญหาการเดาและความบกพร่องในการตอบแบบทดสอบ
- 10) ใช้ในการจัดชั้นเรียนตามระดับความสามารถ (grade-placement tailoring)

ข. ลักษณะเฉพาะของโปรแกรมวิเคราะห์ข้อทดสอบแบบราชที่จะพัฒนาขึ้นใช้

1. มีค่าสถิติที่สำคัญคือ ค่าความยากของข้อทดสอบ (β) ค่าความสามารถของผู้สอบ (θ) ค่าความเชื่อถือในการจำแนกของแบบทดสอบ (Separability Index)
2. มีการทดสอบความสอดคล้องกันระหว่างโค้งลักษณะข้อทดสอบที่คาดหวัง (Expected Item Characteristic Curve) กับ โค้งลักษณะข้อทดสอบที่ปรากฏ (Observed Item Characteristic Curve) ในลักษณะเรียงข้อ (Item Order) และเรียงลำดับความสอดคล้อง (Fit Order)
3. มีแผนผังแสดงการกระจายของข้อทดสอบเรียงลำดับจากข้อที่ง่ายที่สุด ไปยังข้อที่ยากที่สุด

ค. ข้อเสนอแนะในการวิเคราะห์ข้อทดสอบตามแบบราช

1. การวิเคราะห์ข้อทดสอบตามแบบราช สามารถใช้วิเคราะห์ข้อทดสอบของแบบทดสอบวิชาต่างๆ ได้ดี เช่น คณิตศาสตร์ ภูมิศาสตร์ ชีววิทยา และภาษาอังกฤษ แบบทดสอบอาจเป็นแบบทดสอบวินิจฉัย แบบทดสอบชุด (battery test) แบบทดสอบอิงเกณฑ์ แบบทดสอบผลสัมฤทธิ์ และแบบทดสอบเชาวน์ปัญญา หรือความถนัดเฉพาะทาง ซึ่งแบบทดสอบเหล่านี้เป็นแบบทดสอบแบบเลือกตอบ
2. การวิเคราะห์ข้อทดสอบตามแบบราชใช้ได้ดีกับแบบทดสอบอิงเกณฑ์ และแบบทดสอบแบบอิงกลุ่ม แต่โดยปกติแล้วจะใช้ได้ดีกับแบบทดสอบที่มีลักษณะเป็นเอกพจน์และค่อนข้างง่าย
3. ลักษณะของเนื้อหา (test content) และรูปแบบของข้อทดสอบ (item form) ไม่ใช่ลักษณะที่จะกำหนดว่าแบบทดสอบนั้นเหมาะสำหรับที่จะใช้กับแบบราชหรือไม่ เช่น ข้อทดสอบปลายเปิด (open-ended item) และแบบทดสอบถามชนิดลิเคอร์ต (Likert-type) ก็สามารถใช้วิเคราะห์ด้วยแบบราชได้

3. โปรแกรมธนาคารข้อทดสอบ (Item Bank) (สุพัฒน์ สุขมลสันต์, 2539; อุทัย บุญประเสริฐ, 2535)

ก. ความสำคัญของธนาคารข้อทดสอบ

ธนาคารข้อทดสอบ (item bank) หมายถึง ที่รวมของข้อทดสอบ ซึ่งมีไว้เพื่อประโยชน์ในการใช้ข้อทดสอบเหล่านั้นตามวัตถุประสงค์ของการทดสอบในโอกาสต่อไป

ส่วนการจัดเก็บรวบรวม วิธีการดูแลรักษา การดำเนินงานและระบบจัดเก็บ จะมีการดำเนินงานอย่างมีระเบียบสอดคล้องกับวัตถุประสงค์ทางการวัดและประเมินผลการเรียนการสอนมากน้อยเพียงใด

และข้อทดสอบที่จัดเก็บไว้นั้นมีประสิทธิภาพสอดคล้องกับการวัดผลมากน้อยเพียงใดขึ้นอยู่กับความก้าวหน้า และลักษณะการจัดเก็บข้อทดสอบในแต่ละแห่ง โดยทั่วไปแล้วธนาคารข้อทดสอบที่สร้างขึ้นจะช่วยอำนวยความสะดวกในประเด็นต่อไปนี้ คือ

1. ช่วยอำนวยความสะดวกแก่ครูหรืออาจารย์ ในการเลือกข้อทดสอบที่เคยใช้แล้วมาใช้ใหม่ หรือนำมาปรับปรุงให้เหมาะสมยิ่งขึ้น
2. สำหรับธนาคารข้อทดสอบที่มีข้อทดสอบจำแนกตามหมวดหมู่ ตามการวัดการเรียนรู้ ตามกลุ่มเนื้อหาวิชา และตามวัตถุประสงค์ของการทดสอบ ซึ่งครูผู้สอนหรือผู้ที่ทำการทดสอบสามารถเลือกข้อทดสอบต่าง ๆ มาใช้ได้สะดวก และตอบสนองวัตถุประสงค์ของการทดสอบแต่ละครั้ง ได้ดียิ่งขึ้น
3. ช่วยให้ครูหรืออาจารย์ตื่นตัวและให้ความสนใจในการวัดผลประเมินผลการเรียนการสอน เนื่องจากไม่จำเป็นต้องออกข้อทดสอบใหม่ทั้งหมดทุกครั้งที่มีการทดสอบ ผู้ใช้สามารถเลือกข้อทดสอบเก่ามาใช้หรือนำมาปรับปรุงใช้ใหม่ได้โดยง่าย และสามารถพัฒนาการทดสอบที่เป็นอยู่ให้ดียิ่งขึ้น ผลในทางอ้อมก็คือ ธนาคารข้อทดสอบจะมีส่วนช่วยให้ครูหรืออาจารย์ใส่ใจในการตรวจสอบระบบการเรียนการสอนของตน เพื่อหาทางแก้ไข ปรับปรุงและพัฒนาการเรียนการสอนให้ดียิ่งขึ้นได้โดยง่าย โดยพิจารณาจากความสัมพันธ์ระหว่างการทดสอบกับการจัดระบบการเรียนการสอนในเชิงระบบ
4. ช่วยให้ผู้บริหารและผู้รับผิดชอบดูแลงานวิชาการ ได้มีเครื่องมือสำหรับตรวจสอบระดับมาตรฐานทางวิชาการ และพัฒนาคุณภาพการศึกษาได้อย่างมีหลักเกณฑ์ยิ่งขึ้น
5. เป็นบันไดสำคัญที่นำไปสู่การปรับปรุงและพัฒนาการทดสอบและการเรียนการสอนให้ดียิ่งขึ้น มีส่วนช่วยให้ครูหรืออาจารย์ได้เรียนรู้ถึงสิ่งที่สำคัญต่อการเรียนการสอนที่ตนรับผิดชอบอยู่ และสามารถหรือมีโอกาสพัฒนาตนเองในด้านการวัดผลและประเมินผลการเรียนการสอนได้สะดวกยิ่งขึ้น
6. ช่วยลดการสูญเสียแรงงานซ้ำซ้อน ในการจัดทำข้อทดสอบใหม่ทุกครั้งที่มีการทดสอบ

ค. ความสำคัญของธนาคารข้อทดสอบโดยใช้คอมพิวเตอร์ (Computerized Item Bank)

ปัจจุบันเป็นยุคของคอมพิวเตอร์ ดังนั้นหลายสิ่งหลายอย่างจึงมักจะนำเอาคอมพิวเตอร์เข้ามาช่วย เพื่อให้การทำงานเป็นไปได้อย่างรวดเร็ว ถูกต้อง และแม่นยำตรง เช่น การคำนวณที่ซับซ้อน การเก็บรวบรวมข้อมูล และการค้นหาข้อมูล เป็นต้น ในทางการศึกษาได้เริ่มมีการนำคอมพิวเตอร์มาช่วยสอน (computer-assisted instruction) บ้างแล้ว ในด้านการทดสอบนั้นคอมพิวเตอร์มีประโยชน์มากในการทดสอบเฉพาะรายบุคคล (tailored testing หรือ adaptive testing) การสร้างธนาคารคำถาม (Question Bank) และการสร้างธนาคารข้อทดสอบ

การนำความรู้เกี่ยวกับคอมพิวเตอร์มาช่วยในการทำให้ธนาคารข้อทดสอบมีระบบเก็บรวบรวมข้อทดสอบได้เป็นจำนวนมาก ประหยัดเนื้อที่ สามารถเก็บความลับได้เป็นอย่างดี และมีความสะดวกรวดเร็วในการสร้างแบบทดสอบชุดใหม่ตามเกณฑ์ที่ประสงค์

จากหลักการและเหตุผลดังกล่าว การจัดทำธนาคารข้อทดสอบโดยใช้คอมพิวเตอร์จึงมีประโยชน์ดังต่อไปนี้ (สุพรรณ สุกมลสันต์, 2539)

1. ทำให้เนื้อหาของการทดสอบมีความสอดคล้องสัมพันธ์กับจุดมุ่งหมายและเนื้อหาของราย วิชา มากขึ้น และการเรียนการสอนในรายวิชาต่างๆมีความสมบูรณ์ยิ่งขึ้น
2. ทำให้แบบทดสอบมีคุณภาพและน่าเชื่อถือมากยิ่งขึ้น ซึ่งจะทำให้ผลการทดสอบมีความน่าเชื่อถือ มากขึ้นด้วย
3. เป็นการพัฒนาข้อทดสอบให้มีมาตรฐานสูงขึ้น เช่น มีความตรง (validity) และความเที่ยง (reliability) ความไม่มีอคติ (unbiased) หรือการเดา (guessing) ลดน้อยลง เป็นต้น
4. สามารถสร้างแบบทดสอบที่เหมาะสมกับกลุ่มผู้สอบเป้าหมาย (target testees) ได้อย่างรวดเร็ว และถูกต้องตามหลักการทดสอบ ทั้งแบบทดสอบผลสัมฤทธิ์ (achievement test) แบบทดสอบสมรรถภาพ (proficiency test) แบบทดสอบวินิจฉัย (diagnostic test) และแบบทดสอบคัดเลือก (screening test) เป็นต้น
5. สามารถใช้ข้อทดสอบข้อเดียวกันได้หลายๆครั้ง และมีความเหมาะสมกับจุดมุ่งหมายในการ ทดสอบแต่ละครั้ง ได้อย่างรวดเร็ว ซึ่งเป็นการประหยัดทั้งแรงคน การเงิน ได้เป็นอย่างดี ทำให้ข้อทดสอบและแบบทดสอบมีความปลอดภัยจากปัญหาข้อทดสอบที่อาจรั่วไหล
6. ทำให้กระบวนการทดสอบรายวิชาต่างๆ (หรือการทดสอบต่างๆ) มีความพร้อมตลอดเวลา และ สามารถสร้างแบบทดสอบเพื่อการทดสอบได้ทุกเวลาตามที่ผู้บริหารการทดสอบต้องการ
7. ทำให้แบบทดสอบมีลักษณะคู่ขนานกันทั้งในเชิงเนื้อหาและสถิติ (content and statistical parallel test forms) ได้ง่าย เพื่อประโยชน์ในการเทียบคะแนน (score equating) ของแบบทดสอบต่างชุดกัน แต่มี จุดมุ่งหมายเหมือนกัน
8. เป็นการประหยัดเนื้อที่ของสถานที่สำหรับเก็บรักษาแบบทดสอบจำนวนมาก เพื่อเก็บไว้ใช้อีก หรือเพื่อการพัฒนาคุณภาพให้ดีขึ้น
9. เป็นประโยชน์ในการสร้างแบบทดสอบสำหรับการทดสอบรายบุคคลหรือแบบทดสอบ ปรับเปลี่ยน (Adaptive Test) ในอนาคตได้ด้วย

ง. ลักษณะเฉพาะของโปรแกรมธนาคารข้อทดสอบที่จะพัฒนาขึ้นใช้

1. สามารถจัดเก็บข้อทดสอบปรนัยแบบเลือกตอบและแบบอื่นๆ ได้
2. สามารถจัดเก็บแบบทดสอบแบบต่างๆ ทั้งแบบปรนัยและอัตนัยได้
3. สามารถจัดเก็บข้อทดสอบและแบบทดสอบตามเกณฑ์ที่กำหนดไว้ได้
4. สามารถจัดเก็บแฟ้มตัวอักษร(text file) และแฟ้มภาพ (graphic file) ได้
5. สามารถสร้าง ปรับปรุงแก้ไข ลบทิ้ง และเพิ่มเติมข้อทดสอบและแบบทดสอบได้
6. สามารถเลือกข้อทดสอบและแบบทดสอบตามเกณฑ์ที่ต้องการได้
7. สามารถพิมพ์ข้อทดสอบและแบบทดสอบที่เลือกไว้แล้วได้
8. มีระบบตรวจสอบความถูกต้องของข้อมูล

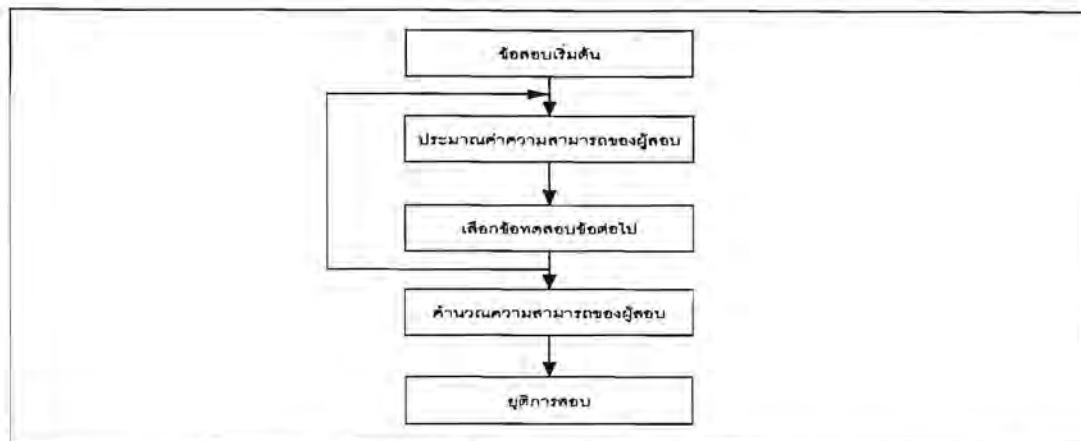
4. โปรแกรมการทดสอบปรับเปลี่ยน (Adaptive Testing)

ก. ความสำคัญของการทดสอบปรับเปลี่ยน

การทดสอบปรับเปลี่ยนคือการทดสอบที่มีการคัดเลือกข้อทดสอบจากธนาคารข้อทดสอบที่ได้ผ่านการวิเคราะห์มาแล้วว่ามีความเหมาะสมกับระดับความสามารถของผู้สอบแต่ละคนในขณะที่สอบ เพื่อให้ได้สารสนเทศสูงสุดสำหรับการประมาณค่าความสามารถที่แท้จริงของผู้สอบ การทดสอบปรับเปลี่ยนด้วยคอมพิวเตอร์(Computerized Adaptive Testing) สามารถจัดข้อทดสอบต่างชุดกันให้เหมาะสมกับความสามารถของผู้สอบแต่ละคนได้ พร้อมทั้งสามารถประมาณค่าความสามารถของผู้สอบได้ด้วย ความคลาดเคลื่อนต่ำ และผลการประมาณค่าสามารถนำมาเปรียบเทียบได้อย่างเที่ยงธรรม การทดสอบในลักษณะนี้จึงมีความยืดหยุ่น ใช้ข้อทดสอบจำนวนน้อย ใช้เวลาในการทดสอบน้อย ทดสอบเวลาใดก็ได้ และทดสอบแล้วทราบผลการทดสอบทันที จึงเป็นการทดสอบที่มีประสิทธิภาพสูง (ศิริชัย กาญจนวาสิ, 2545; สุวัฒน์ สุขมตสันต์, 2539)

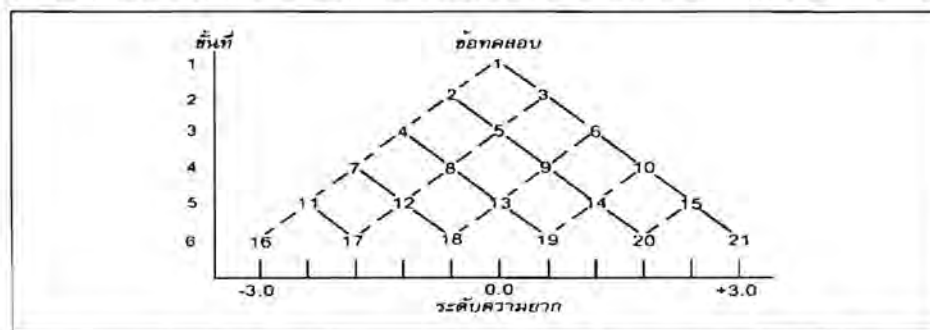
ข. ลักษณะเฉพาะของโปรแกรมการทดสอบปรับเปลี่ยนที่จะพัฒนาขึ้นใช้

1. สามารถทำงานได้ตามหลักการของการทดสอบปรับเปลี่ยน (Principles of Adaptive Testing) กล่าวคือ เมื่อผู้สอบทำข้อทดสอบข้อเริ่มต้น โปรแกรมจะประเมินความสามารถของผู้สอบเบื้องต้น แล้วคัดเลือกข้อทดสอบข้อต่อไปที่มีค่าความเหมาะสมที่จะใช้วัดความสามารถของผู้สอบ แล้วประมาณระดับความสามารถของผู้สอบใหม่ จากนั้นก็จะเลือกข้อทดสอบที่เหมาะสมต่อไป โดยอาศัยหลักการที่ว่า ถ้าการทำข้อทดสอบที่ผ่านมาถูก ข้อถัดไปจะยากขึ้น แต่ถ้าทำข้อทดสอบที่ผ่านผิด ข้อถัดไปจะง่ายลง กระบวนการนี้จะดำเนินการต่อไปเรื่อยๆจนถึงเกณฑ์ที่กำหนดไว้ จนกว่าการทดสอบจะยุติลง แนวคิดดังกล่าวมาสามารถแสดงเป็นแผนภาพได้ดังนี้



ภาพประกอบที่ 4: แนวคิดเกี่ยวกับการทดสอบปรับเปลี่ยนกับความสามารถของผู้สอบ

สำหรับ โปรแกรมทดสอบแบบปรับเปลี่ยนที่จะพัฒนาขึ้นใช้ในการวิจัยครั้งนี้อาศัยแนวคิดจากการทดสอบปรับเปลี่ยนแบบพีระมิดชนิดขนาดขั้นคงที่(constant step size pyramidal model) ได้แก่ การทดสอบปรับเปลี่ยนที่จัดเรียงข้อทดสอบเป็นรูปสามเหลี่ยม โดยมีจำนวนข้อทดสอบในแต่ละขั้นเท่ากับลำดับที่ของขั้น กล่าวคือ ขั้นที่ 1 จะมีข้อทดสอบ 1 ข้อ ขั้นที่ 2 จะมีข้อทดสอบ 2 ข้อ และในขั้นที่ 5 จะมีข้อทดสอบ 5 ข้อ ดังนั้นการทดสอบปรับเปลี่ยนแบบพีระมิดที่มี 6 ขั้น จะมีข้อทดสอบทั้งหมด 21 ข้อ ดังแสดงในภาพต่อไปนี้



ภาพประกอบที่ 5: โครงสร้างของการทดสอบปรับเปลี่ยนแบบพีระมิดชนิดที่มีขนาดคงที่ (สุพัฒน์ สุกมลสันต์, 2539)

จากภาพประกอบที่ 5 แขนงอนเป็นระดับความยากของข้อทดสอบ (β) ที่สัมพันธ์กับข้อทดสอบในการทดสอบปรับเปลี่ยนแบบพีระมิด ซึ่งจะเห็นว่าระดับความยากจะมีระยะห่างกันช่วงละ 0.6 ข้อทดสอบที่อยู่ในแนวตั้งแนวเดียวกันจะมีระดับความยากเท่ากัน และช่วงห่างของระดับความยากของข้อที่อยู่ติดกันภายในขั้นจะมีค่าเท่ากันตลอดคือ 1.2

ดังนั้นในการตอบแบบทดสอบ ผู้สอบจะต้องตอบข้อทดสอบขึ้นละหนึ่งข้อ โดยเริ่มตอบข้อที่อยู่บนยอดสามเหลี่ยมซึ่งมีระดับความยากปานกลางก่อน ถ้าตอบถูกขั้นต่อไปจะแยกไปตามข้อที่ยากขึ้น แต่ถ้าหากตอบผิด ขั้นต่อไปจะตอบข้อที่ง่ายกว่า และกระบวนการทดสอบจะเป็นเช่นนี้จนถึงขั้นสุดท้าย

5. โปรแกรมกำหนดระดับคะแนน (Grade Assignment)

ก. ความสำคัญของการกำหนดระดับคะแนน

1. การประเมินผลการเรียนเป็นหนึ่งในสามของปัจจัยที่สำคัญในการเรียนการสอนรายวิชาต่างๆ นอกเหนือจากวัตถุประสงค์ในการเรียนการสอน และกิจกรรมการเรียนการสอน (Tylor, 1986) ส่วนการกำหนดระดับคะแนน (grading) เป็นกิจกรรมการประเมินผลการเรียนอย่างหนึ่งที่นิยมทั่วไปในสถาบันการศึกษา ดังนั้น ครูผู้สอนควรต้องเรียนรู้วิธีการต่างในการกำหนดระดับคะแนนและดำเนินการให้ถูกต้อง
2. การกำหนดระดับคะแนนทุกวิธีต้องอาศัยพื้นฐานความรู้ที่เกี่ยวข้องกับวิธีการกำหนดระดับคะแนนและการคำนวณ ดังนั้น จึงอาจเกิดการผิดพลาดได้ นักวัดและประเมินผลทางการศึกษาที่มีชื่อเสียงมากคนหนึ่งเคยกล่าวไว้ว่า “การกำหนดระดับคะแนนผิดของครูก่อให้เกิดความเสียหายมากกว่าการกระทำอย่างอื่น” (Terwilliger, 1977)
3. การกำหนดระดับคะแนนมีหลายวิธีแล้วแต่แนวความเชื่อทางการศึกษา วิธีการเรียนการสอน รวมทั้งวิธีวัดและประเมินผลของครูผู้สอน หรือปรัชญาความเชื่อของแต่ละสถานศึกษา ปัจจัยต่างๆ ที่หลากหลายเหล่านี้อาจเป็นเหตุให้การกำหนดระดับคะแนนของผู้สอนผิดพลาดได้ง่าย

ข. ลักษณะเฉพาะของโปรแกรมการกำหนดระดับคะแนนที่จะพัฒนาขึ้นใช้

1. สามารถกำหนดระดับคะแนนได้ 7 วิธี คือ
 - 1.) วิธีใช้ค่าร้อยละที่กำหนดไว้ล่วงหน้า (Fixed Percentage Method)
 - 2.) วิธีใช้ค่าร้อยละ (Percentage Method)
 - 3.) วิธีใช้คะแนนดิบ (Raw Score Method)
 - 4.) วิธีใช้พิสัยของคะแนนดิบ (Ranges of Raw Score Method)
 - 5.) วิธีใช้ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐาน (Mean and S.D. Method)
 - 6.) วิธีใช้ค่าเฉลี่ยและค่าความคลาดเคลื่อนมาตรฐานในการวัด (Mean and SEM Method)
 - 7.) วิธีใช้ค่าคะแนนมาตรฐานที่ (T-Score Method)
2. การกำหนดระดับคะแนนแต่ละวิธีผู้ใช้โปรแกรมสามารถเลือกเป็นแบบ 5 หรือ 8 ระดับคะแนน
3. แต่ละวิธีผู้ใช้โปรแกรมสามารถเลือกให้โปรแกรมพิมพ์คะแนนดิบด้วยหรือไม่ก็ได้

4. โปรแกรมสามารถคิดจำนวนผู้สอบได้แต่ละระดับคะแนน และจำนวนผู้เข้ารับการประเมินได้

ส่วนที่ 5 งานวิจัยที่เกี่ยวข้องกับการสร้างและพัฒนาแบบทดสอบสมิทริภาพทั่วไปทางภาษาอังกฤษ

งานวิจัยที่เกี่ยวข้องกับการสร้างและพัฒนาแบบทดสอบสมิทริภาพทั่วไปทางภาษาอังกฤษเพื่อให้บริการแก่นิสิต นักศึกษา และบุคคลทั่วไปในบริบทของสถาบันภาษามีไม่มากนัก แม้ว่าสถาบันภาษาจะให้บริการการทดสอบแก่กลุ่มบุคคลดังกล่าวแล้วมาเป็นเวลานาน และสร้างแบบ ทดสอบชนิดต่างๆ เพื่อให้บริการแต่ละปีจำนวนหลายชุด ทั้งนี้อาจเป็นเพราะคิดว่าการสร้างและการพัฒนาแบบทดสอบเป็นงานประจำอย่างหนึ่งของสถาบันภาษาแล้ว

การสร้างและพัฒนาแบบทดสอบสมิทริภาพทั่วไปทางภาษาอังกฤษเพื่อให้บริการแก่นิสิต นักศึกษา และบุคคลทั่วไปของสถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัยเริ่มกระทำอย่างจริงจังในลักษณะงานวิจัยเพื่อพัฒนา (research and development) ในปี พ.ศ. 2533 เมื่อฝ่ายวิจัยและพัฒนาของสถาบันภาษา และฝ่ายวิชาการของมหาวิทยาลัยร่วมกันทำการวิจัยและพัฒนาแบบทดสอบวัดสมิทริภาพทั่วไปทางภาษาอังกฤษขึ้น เป็นโครงการวิจัย 3 ปี คือ ปี พ.ศ. 2533-2536 แบบทดสอบที่สร้างขึ้นตามโครงการดังกล่าวมีชื่อว่า แบบทดสอบสมิทริภาพทางภาษาอังกฤษของจุฬาลงกรณ์มหาวิทยาลัย (Chulalongkorn University Test of English Proficiency) หรือที่รู้จักกันทั่วไปในปัจจุบันนี้คือ CU-TEP แบบทดสอบนี้สร้างขึ้นตามลักษณะของแบบทดสอบวัดสมิทริภาพทางภาษาอังกฤษที่เป็นที่รู้จักกันทั่วไป ชื่อ TOEFL (Test of English as a Foreign Language) ของ ETS (Educational Testing Service) แห่งอเมริกา แบบทดสอบ CU-TEP เป็นแบบทดสอบปรนัยชนิดเลือกตอบจำนวน 120 ข้อ ประกอบด้วย ข้อทดสอบการฟังเข้าใจความ 35 ข้อ ข้อทดสอบการอ่านเข้าใจความ 60 ข้อ ข้อทดสอบการเขียน 25 ข้อ และเป็นข้อทดสอบการเขียนแบบเรียงความอีก 3 ข้อ รวมเป็นข้อทดสอบทั้งหมด 123 ข้อ และใช้เวลาสอบ 3 ชั่วโมง จากผลการวิจัยที่เกี่ยวข้องในปี พ.ศ. 2533 (สุพรรณ สุกลมลสันต์ และคณะ, 2533) ปี พ.ศ. 2535 (อัจฉรา วงศ์โสธร และคณะ, 2535) และ ปี พ.ศ. 2536 (สุพรรณ สุกลมลสันต์ และคณะ, 2536) ต่างพบว่าแบบทดสอบ CU-TEP ชุดต่างๆที่สร้างและพัฒนาขึ้นเป็นแบบทดสอบที่มีคุณภาพดีมาก มีความตรงเชิงเนื้อหาและความตรงร่วมสมัยกับแบบทดสอบ TOEFL สูง รวมทั้งมีความเที่ยงสูงอีกด้วย ดังนั้น จากคุณสมบัติที่ดีดังกล่าวทำให้คะแนนผลการทดสอบของแบบทดสอบ CU-TEP สามารถพยากรณ์ผลการทดสอบของแบบทดสอบ TOEFL ได้ใกล้เคียงด้วย จึงทำให้แบบทดสอบ CU-TEP เป็นที่เชื่อถือทั่วไปทั้งในระดับมหาวิทยาลัย และวงการศึกษารวมทั้งวงการธุรกิจบางแห่งภายนอกมหาวิทยาลัยด้วย

อนึ่ง ปัจจุบันนี้สถาบันภาษา เป็นผู้สร้างและพัฒนาแบบทดสอบ CU-TEP ปีละ 1-3 ชุด เพื่อส่งให้แก่ศูนย์ทดสอบแห่งจุฬาลงกรณ์มหาวิทยาลัย ซึ่งรับผิดชอบในการให้บริการทดสอบแก่นิสิตในระดับปริญญาบัณฑิตและบัณฑิตศึกษา รวมทั้งหน่วยงานและบุคคลภายนอกมหาวิทยาลัยด้วย

บทที่ 3

วิธีดำเนินการวิจัย

เพื่อให้ได้ผลการวิจัย ตามวัตถุประสงค์ที่กำหนดไว้แล้ว ผู้วิจัยได้ดำเนินการเป็น 2 ส่วน ดังต่อไปนี้

ส่วนที่ 1 การสร้างและพัฒนาแบบทดสอบ CULI-TEPOIC

ประชากร (population)

ประชากรของการวิจัยครั้งนี้ ได้แก่ นิสิต นักศึกษา และบุคคลภายนอกที่สนใจสอบแบบทดสอบวัดสมรรถภาพทางภาษาอังกฤษเพื่อวิชาชีพและการสื่อสารนานาชาติของสถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย หรือ CULI-TEPOIC [Chulalongkorn University Language Institute Test of English Proficiency for Occupational and International Communication Purposes] ในจุฬาลงกรณ์มหาวิทยาลัยและภายนอก ที่สถาบันภาษา จุฬาลงกรณ์ มหาวิทยาลัยได้ประชาสัมพันธ์ให้ใช้บริการทดสอบโดยไม่ต้องเสียค่าสมัครสอบ และผู้เข้าสอบบางคนมีโอกาสดำเนินการสอบไปสอบแบบทดสอบ TOEIC (Test of English for International Communication) ของ ETS (Educational Testing Service) อีกด้วย ดังนั้นจึงไม่สามารถทราบจำนวนประชากรที่แน่นอนได้

ผลวิจัย (subjects)

ผลวิจัยของการวิจัยครั้งนี้ ได้แก่ นิสิต นักศึกษา และบุคคลภายนอกที่สนใจสอบแบบทดสอบ CULI-TEPOIC ชุดที่ 1, 2 และ 3 บุคคลเหล่านี้ มีความรู้ความสามารถทางภาษาอังกฤษหลากหลาย และสมัครมาสอบเองด้วยวัตถุประสงค์ที่แตกต่างกัน เช่น เพื่อเตรียมตัวสอบชิงทุนไปศึกษาต่อ ณ ต่างประเทศ เพื่อเตรียมตัวสอบคัดเลือกไปทำงาน ณ ต่างประเทศ บางคนสอบเพื่อเตรียมตัวสอบเลื่อนขั้นเงินเดือน หรือตำแหน่ง บางคนมาสอบเพื่อให้ผ่านเกณฑ์ระดับความสามารถทางภาษาตามที่หน่วยงานกำหนดไว้ และบางคนมาสอบเพื่อทดสอบความรู้ทั่วไปทางภาษาอังกฤษของตนเอง เป็นต้น การทดสอบด้วยวัตถุประสงค์ต่าง ๆ เหล่านี้ มีผู้เข้าสอบ จำนวนดังต่อไปนี้

ตารางที่ 10: จำนวนผลวิจัย

ที่	ชื่อแบบทดสอบ	จำนวนผู้สอบ
1	CULI-TEPOIC ชุด 1	116
2	CULI-TEPOIC ชุด 2	124
3	CULI-TEPOIC ชุด 3	120
	รวมทั้งสิ้น	360

เครื่องมือการวิจัย

ก. ลักษณะทั่วไป

เครื่องมือการวิจัยครั้งนี้ได้แก่ แบบทดสอบ CULI-TEPOIC จำนวน 3 ชุด คือ ชุด 1, 2 และ 3 แต่ละชุด เป็นข้อทดสอบปรนัยแบบเลือกคำตอบ 4 ตัวเลือก และมี 2 ส่วน รวมเป็น 100 ข้อ คือ

1. ข้อทดสอบการฟังเข้าใจความ 50 ข้อ
2. ข้อทดสอบการเขียนและอ่านเข้าใจความ 50 ข้อ

ข. การสร้างและการพัฒนา

แบบทดสอบทั้ง 3 ชุด เป็นแบบทดสอบที่จัดสร้างและพัฒนาขึ้น โดยคณะกรรมการสร้างและพัฒนาแบบทดสอบ CULI-TEPOIC ของสถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย ในปี 2546 โดยการอิงแนวคิดและรูปแบบของแบบทดสอบ TOEIC แต่มีวัตถุประสงค์ในการทดสอบแตกต่างกัน กล่าวคือ แบบทดสอบ CULI-TEPOIC สร้างขึ้นเพื่อใช้ทดสอบสมรรถภาพทั่วไปทางภาษาอังกฤษเพื่อวิชาชีพและการสื่อสารนานาชาติ นานาชาติ แต่แบบทดสอบ TOEIC สร้างขึ้นเพื่อใช้ทดสอบสมรรถภาพทั่วไปทางภาษาอังกฤษเพื่อการสื่อสารนานาชาติทั่วไป ขั้นตอนการสร้างและพัฒนาแบบทดสอบดังกล่าวมีรายละเอียดดังต่อไปนี้

1. เสนอแต่งตั้งคณะกรรมการสร้างและพัฒนาแบบทดสอบ CULI-TEPOIC ของสถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย ในปี 2546 จำนวน 7 คน เพื่อสร้างและพัฒนาแบบทดสอบ CULI-TEPOIC จำนวน 3 ชุด
2. คณะกรรมการศึกษาวัตถุประสงค์และรูปแบบของแบบทดสอบ TOEIC เพื่อใช้เป็นแนวทางในการสร้างแบบทดสอบที่ต้องการ
3. คณะกรรมการศึกษาหนังสือและตำราในการสร้างแบบทดสอบภาษาอังกฤษต่างๆ เช่น Alderson (2000), Bachman (1990), Bachman and Palmer (1996), Bloom and Others (1956), Brown (2001), Canale and Swain (1980), Canale (1983), Dunkel (1999), Osterlind (1989) และ Valette (1969) เป็นต้น
4. ผู้วิจัยทำการศึกษาคู่มือและขั้นตอนในการสร้างและพัฒนาแบบทดสอบให้เป็นแบบทดสอบมาตรฐานจากหนังสือ บทความ และตำราต่างๆ เช่น Devine and Yaghlian (2000), Anderson (1972), Zimmerman and Others (1990), Miller and Sundre (2000), Burton and Others (1991) เป็นต้น ซึ่งสรุปได้เป็น 7 ขั้นตอนดังนี้ คือ
 1. สร้างตารางลักษณะข้อทดสอบที่ต้องการเพื่อกำหนดสัดส่วนของข้อทดสอบสำหรับแต่ละวัตถุประสงค์

2. สร้างข้อทดสอบตามลักษณะข้อทดสอบที่กำหนดไว้จำนวนหนึ่ง
 3. ให้ผู้เชี่ยวชาญทางภาษาตรวจสอบความถูกต้องของภาษาและความตรงเชิงเนื้อหาของข้อทดสอบ
 4. ทดลองสอบข้อทดสอบกับกลุ่มตัวอย่างที่เป็นตัวแทนของประชากรเป้าหมายเมื่อวันที่ 20 กรกฎาคม 2546)
 5. ทำการวิเคราะห์ข้อทดสอบรายข้อเพื่อตัดสินใจว่าควรปรับปรุงแก้ไขหรือตัดข้อทดสอบบางข้อทิ้ง
 6. ทดลองสอบข้อทดสอบที่ปรับปรุงแล้วอีกครั้งหนึ่ง
 7. กำหนดหาค่าความเที่ยงและความตรงของแบบทดสอบ
5. ดำเนินการสร้างและพัฒนาแบบทดสอบตามขั้นตอนต่างๆดังกล่าวแล้ว ยกเว้นขั้นตอนที่ 6 เนื่องจาก การวิจัยครั้งนี้พบว่า มีผลวิจัยในการทดสอบแบบทดสอบแต่ละชุดจำนวนมากพอที่จะทำให้ ค่าสถิติในขั้นที่ 5 คงที่เพียงพอ และค่าสถิติต่างๆที่ได้จากการวิเคราะห์ข้อทดสอบรายข้อ รวมทั้ง แบบทดสอบทั้งฉบับสูงมากพอที่จะสามารถนำแบบทดสอบไปใช้ต่อไปได้ นอกจากนี้ยังพบว่ามี ข้อทดสอบที่ควรต้องปรับปรุงแก้ไขจำนวนไม่มากนัก สามารถที่จะปรับปรุงให้ดีขึ้นแล้วนำไปใช้ สำหรับการทดสอบจริงได้ในอนาคต

ดังนั้น ในขั้นที่ 6 จึงเป็นการสุ่มผู้เข้าสอบแบบทดสอบแต่ละชุดประมาณ 25 คน เพื่อไป สอบแบบทดสอบ TOEIC เมื่อวันที่ 23 สิงหาคม 2546 เพื่อนำคะแนนผลการทดสอบมาคำนวณหา ค่าความตรงร่วมสมัย (Concurrent validity) และสร้างตารางพยากรณ์ผลการทดสอบแบบทดสอบ TOEIC จากผลการทดสอบแบบทดสอบ CULI-TEPOIC แทนที่จะทำการทดลองสอบแบบทดสอบ อีกครั้งหนึ่ง

การวิเคราะห์ข้อมูล

1. ทำการวิเคราะห์ข้อทดสอบของแบบทดสอบแต่ละชุด ด้วยวิธีประเพณีนิยม (classical model) เพื่อ หาคุณภาพของแบบทดสอบรายข้อและทั้งฉบับ โดยใช้โปรแกรมคอมพิวเตอร์ชื่อ CTIA/Grading (สุพัฒน์ สุกมลสันต์, 2542: 30-65)
2. ทำการทดสอบหาความเป็นเอกมิติ (unidimensionality) ของแบบทดสอบทั้ง 3 ชุด โดยการ วิเคราะห์องค์ประกอบของแบบทดสอบโดยใช้โปรแกรม SPSS/PC แล้วนำค่า factor loading มาหา ค่าสัมประสิทธิ์สหสัมพันธ์กับค่า Biserial Correlation ซึ่งได้จากขั้นตอนที่ 1 และได้ค่า $r_{xy} = 0.87-0.93$ แสดงว่าแบบทดสอบแต่ละชุดมีความเป็นเอกมิติ (Warm, 1978: 101)

3. วิเคราะห์ข้อทดสอบของแบบทดสอบแต่ละชุด ด้วยวิธีโลจิสต์แบบราช (Logist/Rasch model) เพื่อวิเคราะห์หาค่าความยาก (β) ที่ไม่เปลี่ยนแปลงของข้อทดสอบ ทั้งนี้เพื่อใช้ประกอบการตัดสินใจในการเรียงลำดับข้อทดสอบตามความยากง่ายโดยใช้โปรแกรมคอมพิวเตอร์ชื่อ Rasch (สุพรรณ สุกมลสันต์, 2542: 67-84)
4. นำผลการวิเคราะห์ในขั้นที่ 1 มาศึกษาข้อดี และข้อบกพร่องที่ควรปรับปรุงแก้ไขข้อทดสอบรายข้อให้มีคุณภาพดียิ่งขึ้น โดยอาศัยเกณฑ์การคัดเลือกข้อทดสอบปรนัยแบบเลือกตอบชนิด 4 ตัวเลือกที่เป็นแบบทดสอบสมิทธิภาพทั่วไปดังนี้ (Cheung and Bucat, 2002; DIIA, 2003; Runte, 2001; Jerard, 1995; สุพรรณ สุกมลสันต์, 2542; Ebel, 1972)

ก. ตัวเลือกที่ถูก

1. มีค่าดัชนีความยาก (difficulty index: p) = 0.15 – 0.85
2. มีค่าดัชนีอำนาจจำแนก (discrimination index: d) = > 0.25
3. มีค่าดัชนีความตรงรายข้อ (item validity index: r_{pb}) = > 0.25 หรือ มีนัยสำคัญ

ข. ตัวเลือกที่ผิด

1. มีค่าดัชนีความยาก (difficulty index: p) = 0.15 – 0.85
2. มีค่าดัชนีอำนาจจำแนก (discrimination index: d) = < 0.25 หรือค่าติดลบ
3. มีค่าดัชนีความตรงรายข้อ (item validity index: r_{pb}) = < 0.25 หรือค่าติดลบ หรือไม่มีนัยสำคัญ

ค. ตัวเลือกที่ควรปรับปรุง

1. มีค่าดัชนีความยาก (difficulty index: p) ของข้อที่ถูก < 0.15 หรือ > 0.85
2. ตัวเลือกที่ผิดมีอัตราส่วนของดัชนีความยากแตกต่างจากตัวเลือกผิดอื่นมาก
3. มีค่าดัชนีอำนาจจำแนก (discrimination index: d) < 0.25 หรือมีค่าติดลบ
4. ตัวเลือกที่ผิดมีอัตราส่วนของดัชนีอำนาจจำแนกแตกต่างจากตัวเลือกผิดอื่นมาก
5. มีค่าดัชนีความตรงรายข้อ (item validity index: r_{pb}) < 0.15 หรือมีค่าติดลบ (Jerard, 1995)
6. ตัวเลือกที่ผิดมีอัตราส่วนของดัชนีความตรงรายข้อแตกต่างจากตัวเลือกผิดอื่นมาก
7. ข้อที่ตัวเลือกที่ถูกมีค่าเฉลี่ยของคะแนนเกณฑ์ (Mean criterion score) น้อยกว่าของตัวเลือกที่ผิด

ง. ดัชนีความเที่ยง (reliability index)

1. แบบ KR₂₀ หรือ Cronbach alpha = 0.70 – 0.80 หมายความว่าแบบทดสอบมีความเที่ยงในระดับดี
 2. แบบ KR₂₀ หรือ Cronbach alpha = 0.81 – 0.90 หมายความว่าแบบทดสอบมีความเที่ยงในระดับดีมาก
 3. แบบ KR₂₀ หรือ Cronbach alpha > 0.91 หมายความว่าแบบทดสอบมีความเที่ยงในระดับดีเยี่ยม
 4. แบบ KR₂₀ หรือ Cronbach alpha < 0.60 หมายความว่าแบบทดสอบมีความเที่ยงต่ำ และควรปรับปรุงแก้ไข เว้นแต่มีข้อทดสอบจำนวนน้อย เช่น 10 เป็นต้น
 5. เกณฑ์ทั่วไปของแบบทดสอบปรนัยชนิดเลือกตอบที่มีตัวเลือก 4 ตัวเลือกและมีความยาวมากกว่า 50 ข้อควรมีค่าความเที่ยงแบบ KR₂₀ = >0.80 (Jerard, 1995)
5. คัดเลือกข้อทดสอบที่ดีและข้อที่ควรปรับปรุงแก้ไขจากแบบทดสอบทั้ง 3 ฉบับ เพื่อนำไปแก้ไข สำหรับใช้ในการทดสอบจริงต่อไป

สถิติที่ใช้

1. ความยากง่ายของข้อทดสอบแบบประเพณีนิยม (สุพัตน์ สุขมลสันต์, 2538:18-19)

$$p = \frac{N_a}{N_t}$$

เมื่อ N_a = จำนวนผู้สอบทั้งหมดที่เลือกตอบตัวเลือก i

N_t = จำนวนผู้เข้าสอบทั้งหมด

2. อำนาจจำแนกของข้อทดสอบ (สุพัตน์ สุขมลสันต์, 2538:21)

$$d = \frac{N_u - N_l}{N}$$

เมื่อ d = ค่าอำนาจจำแนก

N_u = จำนวนคนในกลุ่มเก่งจำนวน 27% ที่เลือกตัวเลือก i

N_l = จำนวนคนในกลุ่มอ่อนจำนวน 27% ที่เลือกตัวเลือก i

N = จำนวนคนทั้งหมดในกลุ่ม 27%

3. ความตรงรายข้อของข้อทดสอบ (สุพัฒน์ สุกมลสันต์, 2538:21)

$$r_{pb} = \frac{\overline{x_r} - \overline{x_w}}{S.D._t} \cdot \sqrt{p(1-p)}$$

เมื่อ X_r = ค่าเฉลี่ยของคะแนนผู้สอบที่ตอบข้อทดสอบนั้นถูก

X_w = ค่าเฉลี่ยของคะแนนผู้สอบที่ตอบข้อทดสอบนั้นผิด

SD_t = ค่า SD ของคะแนนรวมของผู้สอบทั้งหมด

p = อัตราส่วนของผู้ที่ตอบข้อทดสอบนั้นถูก

y = ความสูงของแกน Y ในโค้งปกติ ณ จุดที่ค่า P

r_{pb} = ค่าอำนาจจำแนกแบบพอยท์ไบซีเรียล (Point Biserial Correlation)

4. การทดสอบความมีนัยสำคัญของค่าความตรงรายข้อของข้อทดสอบ (สุพัฒน์ สุกมลสันต์, 2538:21)

$$t = \frac{r_{xz}}{\sqrt{(1-r_{xz}^2)/(n-2)}}$$

เมื่อ r_{xz} = ค่าความเที่ยงแบบ r_{pb} หรือ r_{bis}

n = จำนวนผู้สอบทั้งหมดที่ตอบข้อทดสอบนั้น

5. ความเที่ยงของแบบทดสอบ (สุพัฒน์ สุกมลสันต์, 2538:21)

$$KR_{20} = \frac{k}{k-1} \left[1 - \frac{\sum p_i q_i}{S_t^2} \right]$$

เมื่อ k = จำนวนข้อทดสอบ

p_i = อัตราส่วนของผู้ที่ตอบถูกแต่ละข้อ

q_i = $(1 - p_i)$

S_t^2 = ความแปรปรวนของคะแนนรวมของผู้สอบทั้งหมด

KR_{20} = ค่าความเที่ยงของแบบทดสอบชนิดหนึ่ง

6. ความตรงร่วมสมัย(concurrent validity)ของแบบทดสอบ

ความตรงร่วมสมัยคำนวณจากค่าความสัมพันธ์ระหว่างคะแนนผลการทดสอบของแบบทดสอบ 2 ชุด ชุดที่มุ่งทดสอบความรู้ความสามารถอย่างเดียวกัน สามารถคำนวณได้จากสูตรความสัมพันธ์อย่างง่ายแบบเพียร์สัน (Simple Pearson Product-moment Correlation) ดังต่อไปนี้ (Downie and Heath, 1970 : 92)

$$r_{xy} = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}}$$

เมื่อ X = คะแนนสอบแบบทดสอบชุดที่ 1 ของผู้สอบคนที่ i

Y = คะแนนสอบแบบทดสอบชุดที่ 2 ของผู้สอบคนที่ i

N = จำนวนผู้เข้าสอบทั้งหมด

r_{xy} = ค่าความตรงร่วมสมัยของแบบทดสอบ

7. ความยากของข้อทดสอบแบบราช (Rasch model)

การคิดค่าความยากของข้อทดสอบแบบราช มีค่าโดยประมาณดังนี้ (Lord 1980 : 34)

$$\beta_i \approx \frac{\Delta_i}{r_{bis}}$$

เมื่อ Δ_i = ค่าความยากของข้อทดสอบแบบหนึ่ง

r_{bis} = ค่าอำนาจจำแนกของข้อทดสอบแบบไบซีเรียล

β_i = ค่าความยากของข้อทดสอบแบบราช

8. ความสามารถของผู้สอบ (Wright and Mead, 1977 : 8)

$$\theta_i = H + X \ln \left[\frac{r_v}{L - r_v} \right]$$

เมื่อ H = ค่าเฉลี่ยของระดับความยากง่ายทั้งหมด

X = ค่าแก้ไขความคลาดเคลื่อนของ β_i ซึ่งได้จาก

$$X = \sqrt{\frac{1 + W^2}{2.89}}$$

เมื่อ W^2 = ความแปรปรวนของค่าความยาก

L = ความยาวของข้อทดสอบทั้งหมด

r_v = คะแนนของผู้สอบที่มีระดับความสามารถ r

ln = natural log(arithm)

θ_i = ค่าความสามารถของผู้สอบแต่ละคน

ส่วนที่ 2 การสร้างและพัฒนาโปรแกรมคอมพิวเตอร์ที่ใช้ในการพัฒนาแบบทดสอบและการทดสอบ

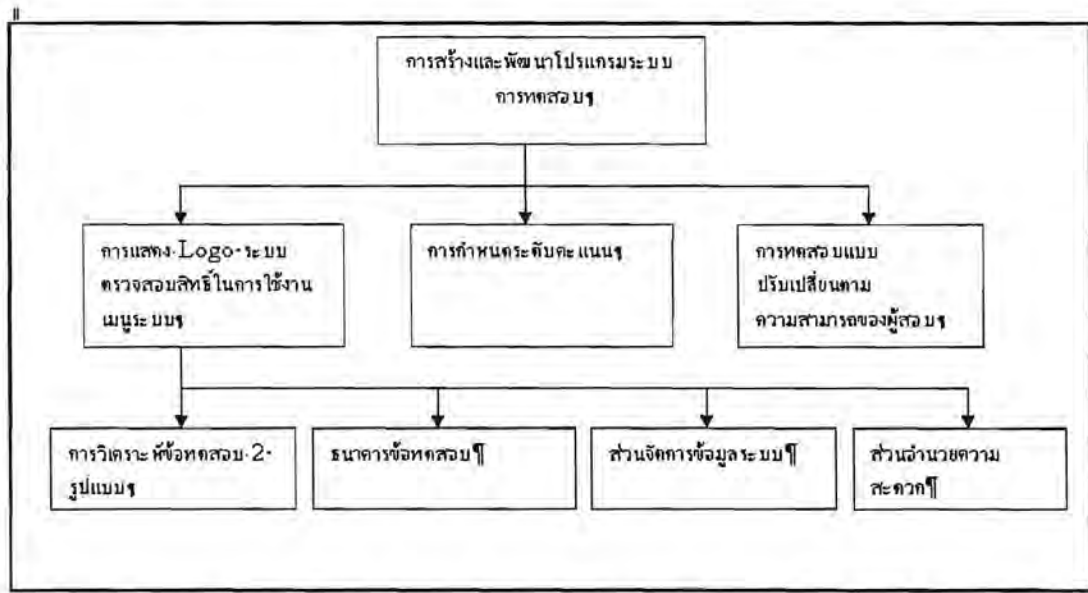
1. ดำเนินการจัดหาและจัดจ้างผู้มีความชำนาญในการเขียนโปรแกรมคอมพิวเตอร์ด้วยภาษา Visual BASIC เพื่อใช้กับระบบ Windows และมีความเข้าใจภาษา Quick BASIC, FORTRAN และ Fox Pro เป็นอย่างดี เนื่องจาก โปรแกรมต่างๆที่ต้องการพัฒนานั้นเดิมผู้วิจัยเขียนด้วยภาษาดังกล่าวแต่ใช้งานได้เฉพาะในระบบ DOS (Disk Operation System) เท่านั้น
2. อธิบายให้ผู้เขียนโปรแกรมเข้าใจการทำงานของทั้ง 5 โปรแกรมที่ผู้วิจัยต้องการพัฒนาให้ใช้ได้กับระบบ Windows รวมทั้งความต้องการให้ผู้เขียนโปรแกรมเขียนด้วยภาษา Visual BASIC เพราะโครงสร้างของภาษาดังกล่าวคล้ายคลึงกันกับของภาษา Quick BASIC และ FORTRAN จะทำให้ผู้เขียนโปรแกรมทำงานได้รวดเร็วและถูกต้องยิ่งขึ้น ส่วนระบบฐานข้อมูลให้ใช้ภาษา Microsoft Access แทนภาษา Fox Pro หนึ่งโปรแกรมทั้ง 5 โปรแกรมได้แก่
 1. โปรแกรมวิเคราะห์ข้อทดสอบแบบประเพณีนิยม (Classical Model)
 2. โปรแกรมวิเคราะห์ข้อทดสอบแบบราช (Rasch Model)
 3. โปรแกรมธนาคารข้อทดสอบ (Item Banking)
 4. โปรแกรมการทดสอบแบบปรับเปลี่ยน (Adaptive Testing) และ
 5. โปรแกรมกำหนดระดับคะแนน (Grade Assignment)
3. มอบรหัสต้นฉบับ (source codes) ทั้งที่เป็นสำเนาอิเล็กทรอนิกส์ (soft copy) และสำเนาแสดงผล (hard copy) ให้กับผู้เขียนโปรแกรม เพื่อให้ดำเนินการพัฒนาตามที่ได้ตกลงกันไว้
4. ตรวจสอบการทำงานของโปรแกรมที่พัฒนาขึ้นเป็นระยะๆ และให้คำแนะนำแก่ผู้เขียนโปรแกรมเป็นระยะๆ เพื่อให้ผลงานเป็นไปตามที่ต้องการและถูกต้องมากที่สุด
5. เมื่อโปรแกรมเสร็จสมบูรณ์แล้ว ผู้วิจัยทำการตรวจสอบการทำงานของโปรแกรมใหม่ทั้ง 5 โปรแกรม โดยการเปรียบเทียบกับผลการทำงานของโปรแกรมเดิมว่าเหมือนกันหรือไม่

รายละเอียดเกี่ยวกับการสร้างและพัฒนาโปรแกรมคอมพิวเตอร์ทั้ง 5 โปรแกรม

1. การสร้างและพัฒนาโปรแกรมระบบการทดสอบ

ดังได้กล่าวมาแล้วในบทที่ 1 ว่า โปรแกรมที่สร้างและพัฒนาขึ้นในงานวิจัยนี้มีทั้งหมด 5 โปรแกรม ซึ่งในงานวิจัยนี้เรียกชื่อรวมทั้งหมดว่า “โปรแกรมชุดเพื่อการทดสอบและกำหนดระดับคะแนนด้วยคอมพิวเตอร์” หรือเรียกสั้นๆว่า CTG Package (Computerized Testing and Grading Package Program) โปรแกรมชุดดังกล่าวนี้ประกอบด้วยโปรแกรมที่สำคัญดังนี้

ผู้พัฒนาโปรแกรมเขียน โปรแกรมโดยใช้ภาษา Visual Basic รุ่น 6.0 ในการสร้างหน้าจอแบบฟอร์ม เพื่อเป็นส่วนติดต่อกับผู้ใช้ (User Interface) และใช้โปรแกรมจัดการฐานข้อมูล Microsoft Access 2000 เป็น ส่วนจัดเก็บของระบบงานทั้งหมด จากขั้นตอนการออกแบบ โครงสร้างของโปรแกรม สามารถแสดงผัง โครงสร้างของชุดคำสั่ง ได้ดังภาพประกอบที่ 6



ภาพประกอบที่ 6: โครงสร้างของโปรแกรมในขั้นตอนการพัฒนาโปรแกรม

จากภาพประกอบที่ 6 สามารถอธิบายในแต่ละส่วนของชุดคำสั่ง ได้ดังนี้

1.) ส่วนแสดง Logo ระบบ ตรวจสอบสิทธิการใช้งาน และเมนูระบบ

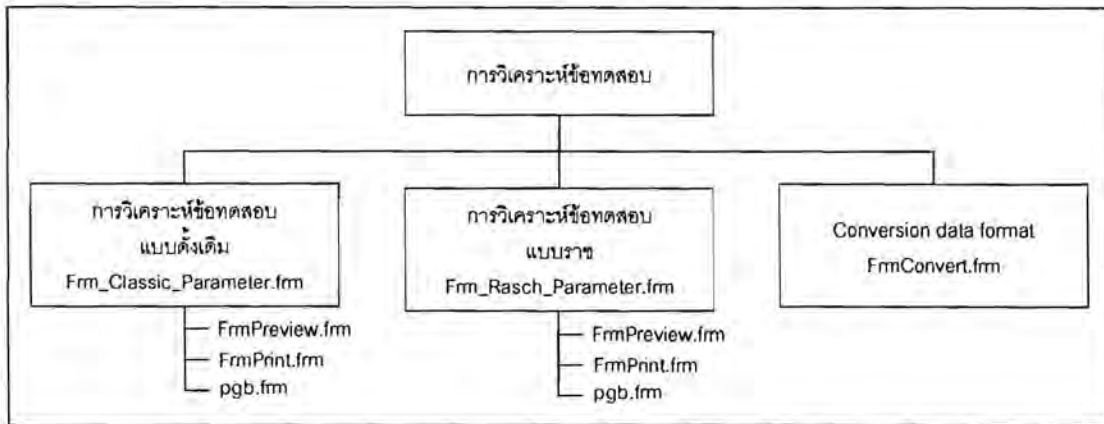
ส่วนนี้แสดง Logo เพื่อตรวจสอบสิทธิการใช้งาน และเมนูหลักของระบบ ซึ่งเป็นส่วน เชื่อมโยงไปยังการทำงานในส่วนต่างๆ ได้แก่ ส่วนการวิเคราะห์ข้อทดสอบ การจัดเก็บข้อทดสอบ ส่วนจัดการ ข้อมูลระบบ และส่วนอำนวยความสะดวกต่างๆ สามารถแสดงรายละเอียดของโปรแกรม ดังตารางต่อไปนี้

ตารางที่ 14: ส่วนแสดง Logo ระบบ ตรวจสอบสิทธิการใช้งาน และเมนูระบบ

ลำดับที่	ชื่อ	หน้าที่การทำงาน
1	FrmSplash.frm	แสดง Logo ของระบบ
2	frm_user_pass.frm	ตรวจสอบสิทธิการใช้งานระบบ
3	MDIForm1.frm	แสดงเมนูหลักของระบบ

2.) ส่วนการวิเคราะห์ข้อทดสอบ

เป็นส่วนของการวิเคราะห์ข้อทดสอบ โดยสามารถแบ่งการวิเคราะห์ข้อทดสอบได้ 2 แบบ ได้แก่ การวิเคราะห์ข้อทดสอบแบบประเพณีนิยม และการวิเคราะห์ข้อทดสอบแบบราช โดยมีโครงสร้างของโปรแกรม ดังภาพประกอบที่ 7 ต่อไปนี้



ภาพประกอบที่ 7: โครงสร้างของโปรแกรมการวิเคราะห์ข้อทดสอบ

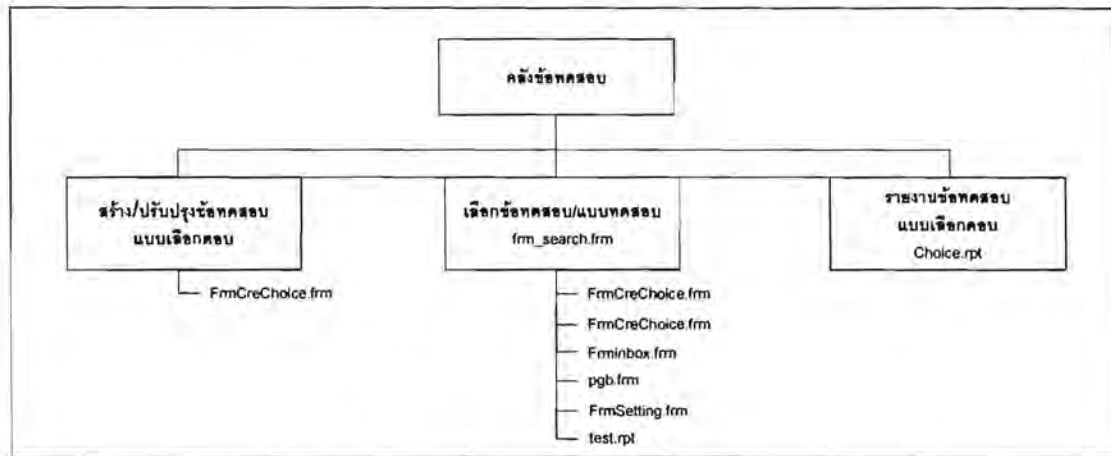
จากภาพประกอบที่ 7 โปรแกรมที่ใช้ในการวิเคราะห์ข้อทดสอบ ประกอบด้วยไฟล์ระบบงาน 5 ไฟล์ โดยมีรายละเอียดของไฟล์ ดังแสดงในตารางที่ 15

ตารางที่ 15: รายละเอียดของไฟล์และการวิเคราะห์ข้อทดสอบ

ลำดับที่ 1	ชื่อ	หน้าที่การทำงาน
1	Frm_Classic_Parameter.frm	การวิเคราะห์ข้อทดสอบแบบประเพณีนิยม
2	Frm_Rasch_Parameter.frm	การวิเคราะห์ข้อทดสอบแบบราช
3	frmPreview.frm	แสดงผลการวิเคราะห์ข้อทดสอบ
4	frmPrint.frm	กำหนดรายละเอียดการพิมพ์
5	pgb.frm	แสดงสถานะการประมวลผล
6	FrmConvert.frm	แปลงรูปแบบข้อมูลนำเข้าจาก Classical ไปเป็น Rasch

3. ส่วนธนาคารข้อทดสอบ

ส่วนนี้ใช้เพื่อการสร้างข้อทดสอบปรนัยแบบเลือกตอบและแบบทดสอบต่างๆด้วย รวมทั้งการปรับปรุงแก้ไข การค้นหา เลือกข้อทดสอบมาใช้งาน การดูผลการทดสอบของการทดสอบแบบปรับเปลี่ยนตามความสามารถของผู้สอบ และรายงานข้อทดสอบที่เก็บอยู่ในธนาคารข้อทดสอบ ส่วนนี้มีโครงสร้างของโปรแกรมดังแสดงในภาพประกอบที่ 8 ต่อไปนี้



ภาพประกอบที่ 8: โครงสร้างของโปรแกรมธนาคารข้อทดสอบ

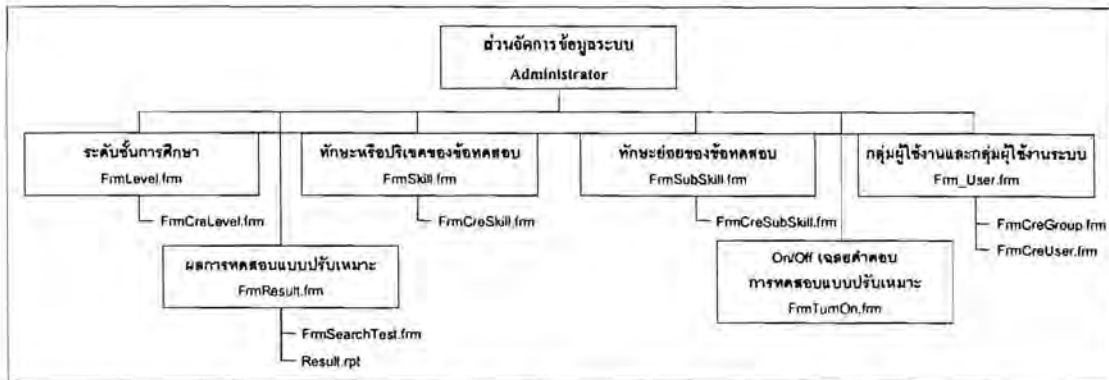
จากภาพประกอบที่ 8 แสดงให้เห็นว่า โปรแกรมที่อยู่ในส่วนของธนาคารข้อทดสอบ ประกอบด้วยไฟล์ระบบงาน 12 ไฟล์ โดยมีรายละเอียดของไฟล์ ดังแสดงในตารางที่ 16 ต่อไปนี้

ตารางที่ 16: รายละเอียดของไฟล์ ในส่วนธนาคารข้อทดสอบ

ลำดับที่	ชื่อ	หน้าที่การทำงาน
1	Frm_Choice.frm	แสดงรายละเอียดข้อทดสอบแบบเลือกตอบ
2	FmCreChoice.frm	สร้างและแก้ไขข้อทดสอบ
3	Frm_Search.frm	แสดงรายละเอียดข้อทดสอบจากการสุ่ม
4	FrmSearch.frm	กำหนดเงื่อนไขการค้นหาข้อทดสอบ
5	fminbox.frm	สุ่มและจัดเรียงข้อทดสอบ
6	pgb.frm	แสดงสถานะการประมวลผล
7	FrmSetting.frm	กำหนดหัวกระดาษและท้ายกระดาษ
8	test.rpt	รายงานข้อทดสอบจากการสุ่ม
9	Choice.rpt	รายงานข้อทดสอบในธนาคารข้อทดสอบ

4. ส่วนจัดการข้อมูลระบบ

โปรแกรมส่วนนี้ทำหน้าที่นำเข้าข้อมูลที่ระบบจำเป็นต้องใช้ ได้แก่ ข้อมูลระดับชั้นการศึกษา ทักษะหรือปริเขตของข้อทดสอบ ทักษะย่อยของข้อทดสอบ การเพิ่ม/ลบกลุ่มผู้ใช้งาน การเพิ่ม/แก้ไขผู้ใช้งาน ระบบ ผลการทดสอบแบบปรับเปลี่ยน และการให้หรือไม่ให้คำตอบ (On/Off) ของ การทดสอบแบบปรับเปลี่ยน ส่วนนี้มีโครงสร้างของโปรแกรมดังแสดงในภาพประกอบที่ 9



ภาพประกอบที่ 9: โครงสร้างของโปรแกรมส่วนจัดการข้อมูลระบบ

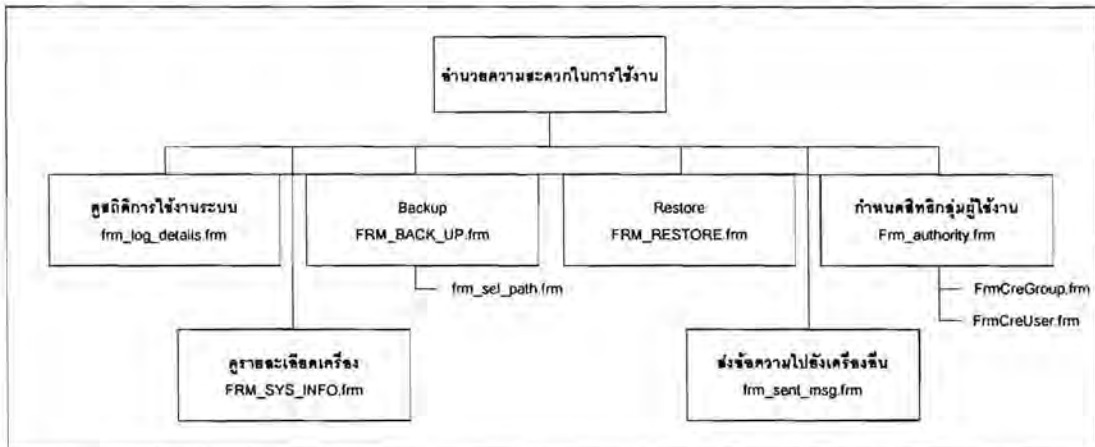
จากภาพประกอบที่ 9 จะเห็นได้ว่า โปรแกรมที่อยู่ในส่วนธนาคารข้อทดสอบ ประกอบด้วย ไฟล์ระบบงาน 13 ไฟล์ ซึ่งมีรายละเอียดของไฟล์ ดังแสดงไว้ในตารางที่ 17 ต่อไปนี้

ตารางที่ 17: รายละเอียดของไฟล์ ในส่วนของการจัดการข้อมูลระบบ

ลำดับที่	ชื่อ	หน้าที่การทำงาน
1	FmLevel.frm	แสดงรายละเอียดระดับชั้นการศึกษา
2	FmSkill.frm	แสดงรายละเอียดทักษะหรือปริเขตของข้อทดสอบ
3	FmSubSkill.frm	แสดงรายละเอียดทักษะย่อยของข้อทดสอบ
4	FmUser.frm	แสดงรายละเอียดผู้ใช้งานระบบ
5	FmCreLevel.frm	เพิ่ม/แก้ไข ระดับชั้นการศึกษา
6	FmCreSkill.frm	เพิ่ม/แก้ไข ทักษะหรือปริเขตของข้อทดสอบ
7	FmCreSubSkill.frm	เพิ่ม/แก้ไข ทักษะย่อยของข้อทดสอบ
8	FmCreGroup.frm	เพิ่มกลุ่มผู้ใช้งาน
9	FmCreUser.frm	เพิ่ม/แก้ไขผู้ใช้งาน
10	FmResult.frm	แสดงรายละเอียดผลการทดสอบแบบปรับเปลี่ยน
11	FmSearchTest.frm	กำหนดเงื่อนไขการค้นหาผลการทดสอบแบบปรับเปลี่ยน
12	Result.rpt	รายงานผลการทดสอบแบบปรับเปลี่ยน
13	FmTurnOn.frm	On/Off เฉลยคำตอบการทดสอบแบบปรับเปลี่ยน

5. ส่วนอำนวยความสะดวก

โปรแกรมส่วนนี้ทำหน้าที่เก็บบันทึกค่าสถิติการเข้ามาใช้งานระบบ การสำรองข้อมูล การนำข้อมูลกลับมาใช้ใหม่ และการกำหนดสิทธิการใช้งานระบบให้กับกลุ่มผู้ใช้งาน นอกจากนี้โปรแกรมยังส่งข้อมูลไปยังเครื่องอื่นที่อยู่ภายในเครือข่ายด้วย และมีโครงสร้างของ โปรแกรมดังแสดงในภาพประกอบที่ 10 ต่อไปนี้



ภาพประกอบที่ 10: โครงสร้างส่วนของโปรแกรมอำนวยความสะดวก

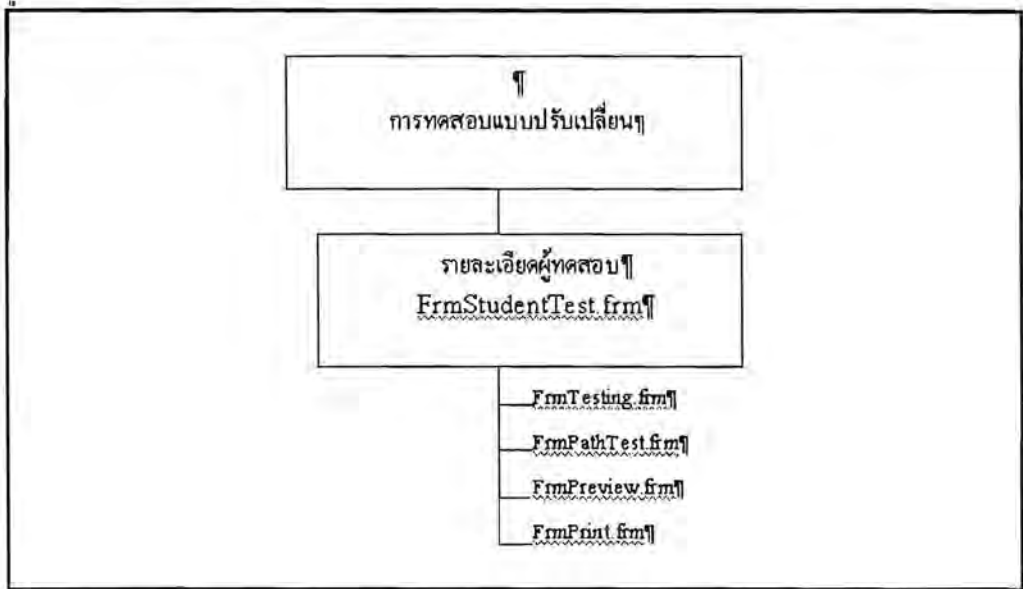
จากภาพประกอบที่ 10 จะเห็นได้ว่าโปรแกรมส่วนนี้ประกอบด้วยไฟล์ระบบงาน 8 ไฟล์ โดยมีรายละเอียดของไฟล์ ดังแสดงในตารางที่ 18

ตารางที่ 18: รายละเอียดของไฟล์ ในส่วนของการอำนวยความสะดวก

ลำดับที่	ชื่อ	หน้าที่การทำงาน
1	Frm_log_details.frm	คู่มือการใช้งานผู้ใช้
2	FRM_BACK_UP.frm	ทำการสำรองข้อมูล
3	Frm_Restore	นำข้อมูลที่สำรองกลับมาใช้ใหม่
4	Frm_authority.frm	กำหนดสิทธิการใช้งานให้กลุ่ม
5	FrmCreGroup.frm	เพิ่ม กลุ่มผู้ใช้งานระบบ
6	FrmCreUser.frm	เพิ่ม/แก้ไข ผู้ใช้งานระบบ
7	FRM_SYS_INFO.frm	ดูรายละเอียดเครื่องคอมพิวเตอร์
8	FrmCreUser.frm	เพิ่ม/แก้ไขผู้ใช้งาน

6. ส่วนการทดสอบแบบปรับเปลี่ยนตามความสามารถของผู้สอบ

โปรแกรมส่วนนี้มีโครงสร้างของโปรแกรมดังแสดงในภาพประกอบที่ 11



ภาพประกอบที่ 11: แสดงโครงสร้าง ส่วนของโปรแกรมการทดสอบแบบปรับเปลี่ยน

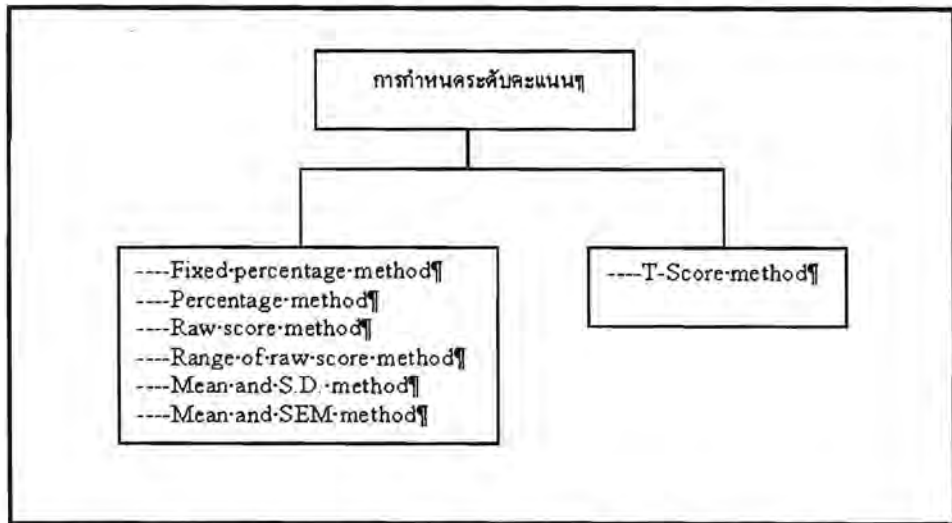
จากภาพประกอบที่ 11 แสดงให้เห็นว่า โปรแกรมที่อยู่ในส่วนของ โปรแกรมการทดสอบแบบปรับเปลี่ยนตามความสามารถของผู้สอบ ประกอบด้วยไฟล์ระบบงาน 8 ไฟล์ โดยมีรายละเอียดของไฟล์ ดังแสดงในตารางที่ 19 ต่อไปนี้

ตารางที่ 19: รายละเอียดของไฟล์ ในส่วนของโปรแกรมการทดสอบแบบปรับเปลี่ยน

ลำดับที่	ชื่อ	หน้าที่การทำงาน
1	FrmStudentTest.frm	กำหนดรายละเอียดผู้เข้ารับการทดสอบ
2	FrmDomainTest.frm	กำหนดประเภทการทดสอบ
3	FrmPathTest.frm	แสดงวิธีการตอบ
4	FrmPreview.frm	แสดงผลการทดสอบ
5	FrmPrint.frm	กำหนดรายละเอียดการพิมพ์

7. ส่วนการกำหนดระดับคะแนน

โปรแกรมส่วนนี้มีโครงสร้างของโปรแกรมดังแสดงในภาพประกอบที่ 12



ภาพประกอบที่ 12: โครงสร้าง ส่วนของ โปรแกรมกำหนดระดับคะแนน

จากภาพประกอบที่ 12 แสดงให้เห็นว่า โปรแกรมส่วนนี้ไม่ต้องอาศัยไฟล์เพิ่มเติมเนื่องจาก โปรแกรมใช้ไฟล์ที่มีอยู่แล้วจากการวิเคราะห์ข้อทดสอบแบบปรับเปลี่ยนสำหรับการกำหนดระดับคะแนน 6 วิธีทางด้านซ้ายมือ ส่วนอีกวิธีหนึ่งนั้นผู้ใช้ต้องป้อนข้อมูลเข้าไปในโปรแกรมเองในภายหลัง

2. การทดสอบการทำงานของโปรแกรมที่สร้างและพัฒนาขึ้น

ก. การเตรียมสภาพแวดล้อมด้านฮาร์ดแวร์และซอฟต์แวร์

ในการทำงานของโปรแกรมการสร้างและพัฒนาโปรแกรมระบบการทดสอบ จำเป็นต้องอาศัย ฮาร์ดแวร์และซอฟต์แวร์ที่ ในการทำงานดังนี้ คือ

1) อุปกรณ์ฮาร์ดแวร์ของแม่ข่าย

- ก. เครื่องคอมพิวเตอร์ หน่วยประมวลผลกลางตั้งแต่ 586 ขึ้นไป
- ข. มีหน่วยความจำหลัก ไม่น้อยกว่า 64 เมกะไบต์
- ค. มีฮาร์ดดิสก์ ความจุไม่น้อยกว่า 2 กิกะไบต์

2) สภาพแวดล้อมด้านซอฟต์แวร์ของแม่ข่าย

- ก. ระบบปฏิบัติการ โปรแกรมวินโดว์ รุ่น 2000 ขึ้นไป
- ข. มีโปรแกรมจัดการระบบฐานข้อมูลแบบสัมพันธ์ของโปรแกรม Microsoft Access 2000

3) อุปกรณ์ฮาร์ดแวร์ของเครื่องลูกข่าย

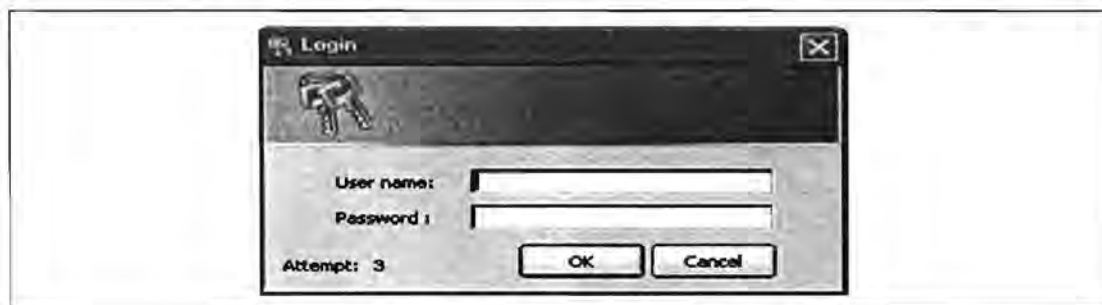
- ก. เครื่องคอมพิวเตอร์ หน่วยประมวลผลกลางตั้งแต่ 586 ขึ้นไป
- ข. มีหน่วยความจำหลักไม่น้อยกว่า 64 เมกะไบต์
- ค. มีฮาร์ดดิสก์ ความจุไม่น้อยกว่า 2 จิกะไบต์

4) สภาพแวดล้อมด้านซอฟต์แวร์ของเครื่องลูกข่าย

- ก. ระบบปฏิบัติการ โปรแกรมวินโดว์ 98 เป็นอย่างน้อย
- ข. มีโปรแกรมการสร้างและพัฒนาโปรแกรมระบบการทดสอบ (CT.exe และ CCAT.exe)

ข. การตรวจสอบสิทธิ์การเข้าใช้งานระบบ CTG Package

การตรวจสอบสิทธิ์การเข้าใช้งานระบบทำได้โดยการเรียกโปรแกรม CTG.exe ขึ้นมาทำงาน ด้วยการคลิกสองครั้ง (double click) ที่ Icon ชื่อ CTG Package แล้วเครื่องคอมพิวเตอร์จะแสดง Logo ของโปรแกรม แล้วระบบจะแสดงหน้าจอให้ป้อนรหัสผู้ใช้งานและ รหัสผ่าน ดังแสดงในภาพประกอบที่ 13



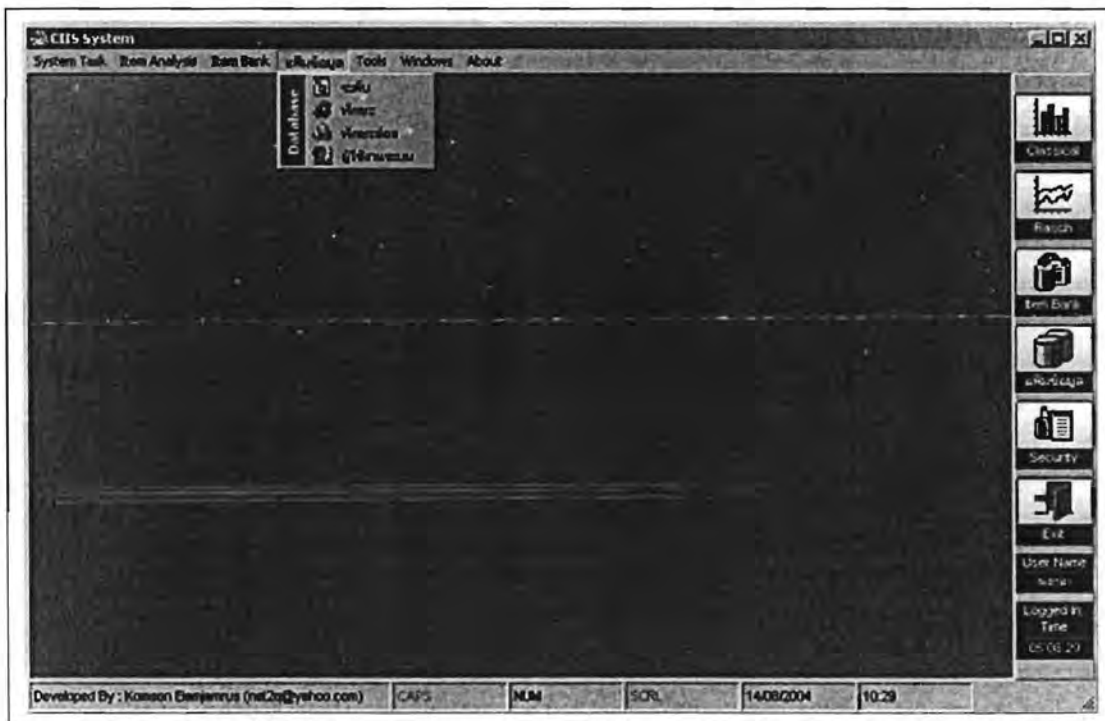
ภาพประกอบที่ 13: แสดงหน้าจอ ให้ป้อนรหัสผู้ใช้งาน และรหัสผ่าน เพื่อตรวจสอบสิทธิ์

โปรแกรมจะให้ผู้ใช้ระบบป้อนรหัสผู้ใช้และรหัสผ่านได้ไม่เกิน 3 ครั้ง จะเตือนหากว่าผู้ใช้ป้อนรหัสเหล่านี้ผิด ดังแสดงไว้ในภาพประกอบที่ 14 และจะออกจากการทำงานทันทีหากว่าผู้ใช้ป้อนรหัสผู้ใช้และรหัสผ่านผิดเกินกว่าจำนวนครั้งที่กำหนดไว้

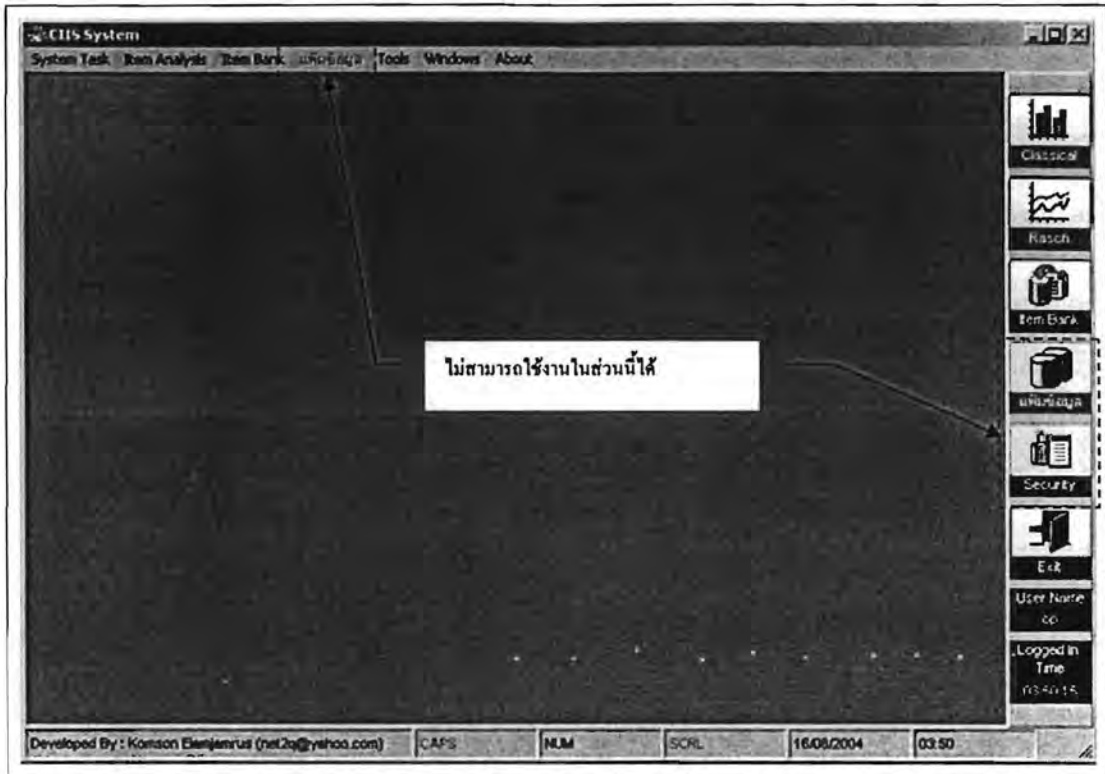


ภาพประกอบที่ 14: แสดงข้อความกรณีป้อนรหัสผู้ใช้หรือรหัสผ่านผิด

เมื่อเข้าสู่ระบบได้แล้วให้ป้อนรหัสผู้ใช้และรหัสผ่านที่เป็นของผู้ดูแลระบบ เพื่อจะดูสิทธิการใช้งานในแต่ละเมนู ทั้งที่เป็นเมนูข้อความและเมนูไอคอน (icon menu) หากว่ารหัสเป็นของผู้ดูแลระบบ ผู้ใช้สามารถใช้โปรแกรมได้ทุกเมนู แต่หากเป็นรหัสของผู้ใช้ทั่วไปจะมีสิทธิใช้โปรแกรมได้เพียงบางเมนูเท่านั้น ดังรายละเอียดแสดงไว้ในภาพประกอบที่ 15-16



ภาพประกอบที่ 15: หน้าจอ เมื่อเข้าใช้งานด้วยรหัสของผู้ดูแลระบบ

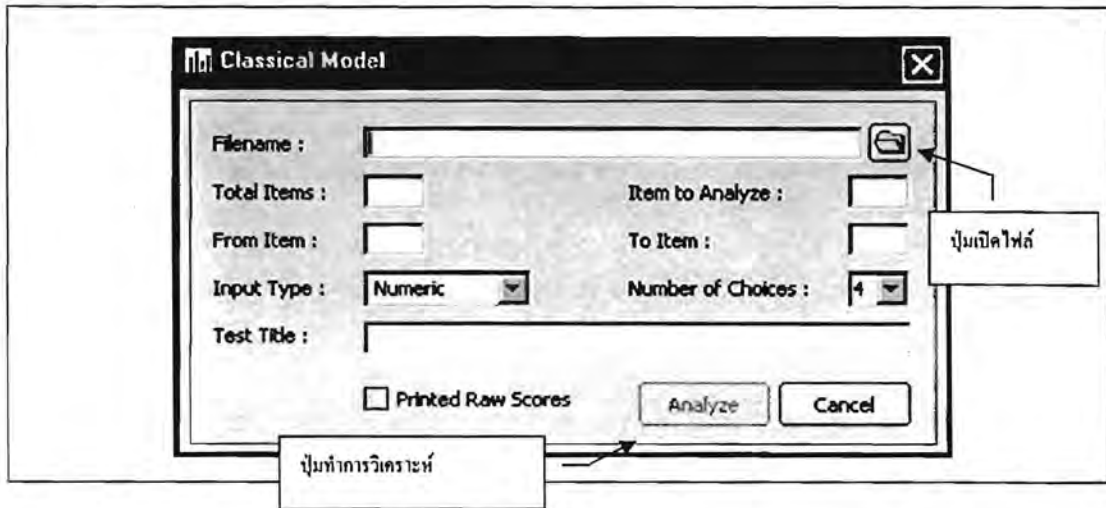


ภาพประกอบที่ 16: หน้าจอ เมื่อเข้าใช้งานด้วยรหัสของผู้ของใช้ทั่วไป

ค. การทดสอบการใช้โปรแกรมที่สร้างและพัฒนาขึ้น

1. การใช้โปรแกรมวิเคราะห์ข้อทดสอบรายข้อแบบประเพณีนิยม (Classical Model)

หน้าจอของโปรแกรมวิเคราะห์ข้อทดสอบแบบประเพณีนิยมมีลักษณะดังที่แสดงไว้ในภาพประกอบที่ 17 กล่าวคือประกอบด้วยปุ่มเปิดไฟล์เพื่อการนำเข้าข้อมูล ฟิลด์แสดงรายละเอียดของข้อมูลที่จะทำการวิเคราะห์ และปุ่ม “Analyze” ซึ่งเป็นปุ่มที่สั่งให้โปรแกรมเริ่มทำการวิเคราะห์ข้อทดสอบปุ่มนี้จะไม่ทำงานจนกว่าจะมีข้อมูลนำเข้า (input data) ที่มีการจัดรูปแบบ (format) ที่ถูกต้อง

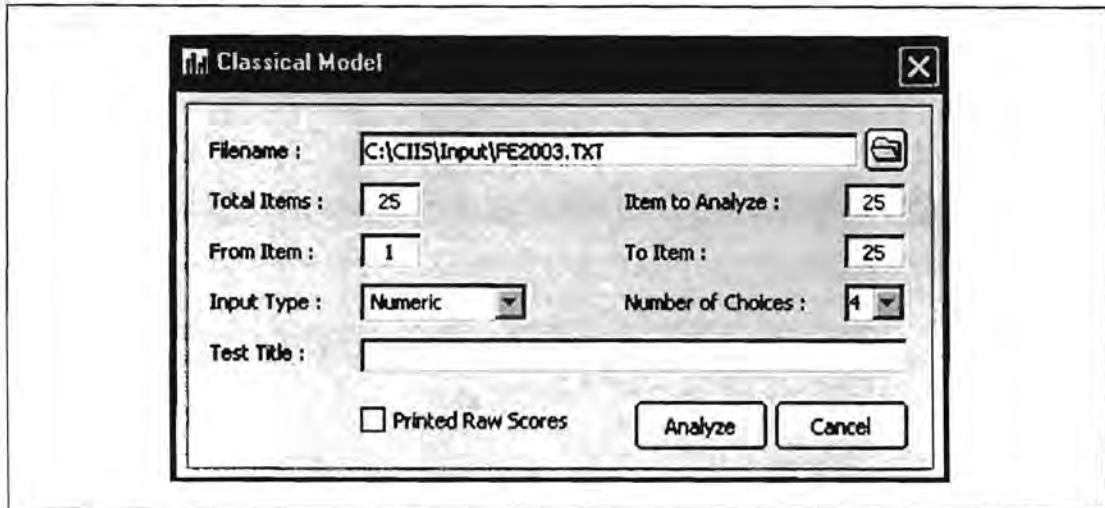


ภาพประกอบที่ 17: หน้าจอ นำเข้าข้อมูลที่จะทำการวิเคราะห์ข้อทดสอบ

ในกรณีที่ข้อมูลนำเข้ามีรูปแบบไม่ถูกต้อง โปรแกรมจะมีข้อความเตือน ดังแสดงไว้ในภาพประกอบที่ 18 ข้างล่างนี้



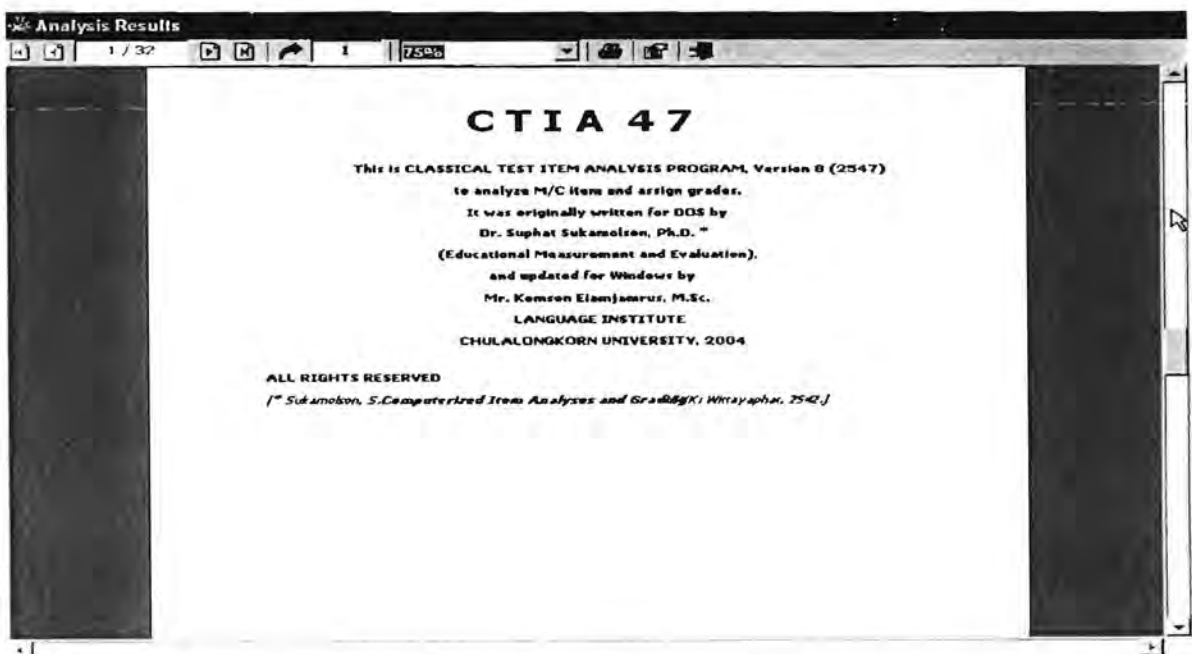
ภาพประกอบที่ 18: แสดงข้อความเตือนเมื่อมุลนำเข้ามีรูปแบบไม่ถูกต้อง



ภาพประกอบที่ 19: หน้าจอ เมื่อข้อมูลนำเข้ามีรูปแบบที่ถูกต้อง

ในกรณีที่ข้อมูลนำเข้ามีรูปแบบที่ถูกต้อง โปรแกรมจะแสดงรายละเอียดของไฟล์นั้นๆ และปุ่ม “Analyze” จะมีสีเข้มขึ้นเพื่อแสดงว่าสามารถใช้งานได้ ดังภาพประกอบที่ 19

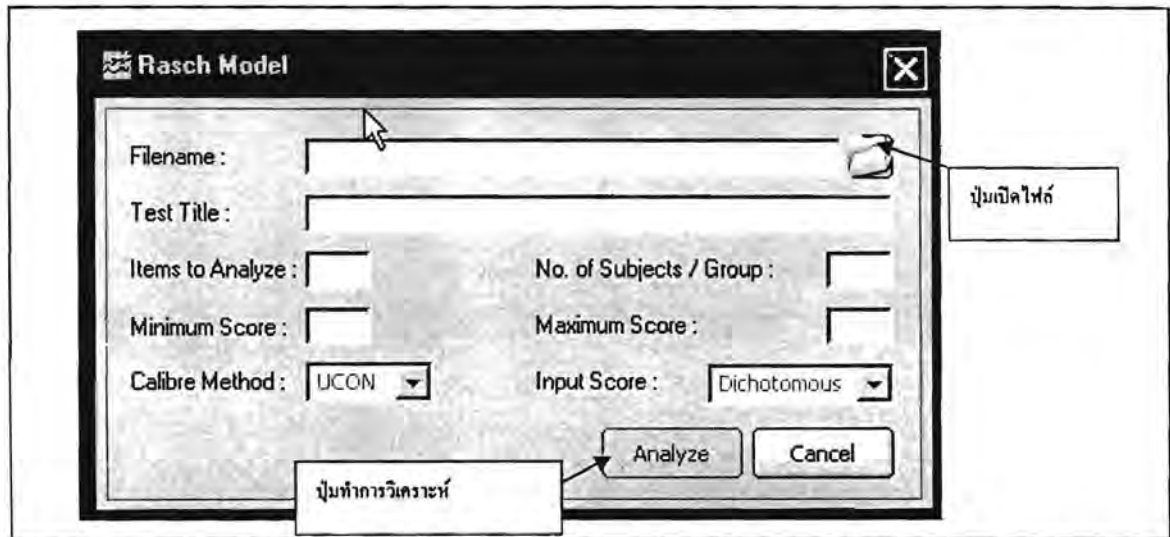
อนึ่ง ที่ฟิลด์ชื่อ “Test Title” ผู้ใช้สามารถใส่ชื่อของแบบทดสอบที่นำมาวิเคราะห์ได้ และชื่อนี้จะไปปรากฏอยู่ที่ไฟล์ของผลการวิเคราะห์ (output file) เมื่อใส่ข้อมูลครบถ้วนแล้วให้คลิกที่ปุ่ม “Analyze” เพื่อทำการวิเคราะห์ข้อทดสอบต่อไป แล้วโปรแกรมจะรายงานผลการวิเคราะห์ดังแสดงไว้ในภาพประกอบที่ 20



ภาพประกอบที่ 20: หน้าจอ แสดงรายงานผลการวิเคราะห์ข้อทดสอบ

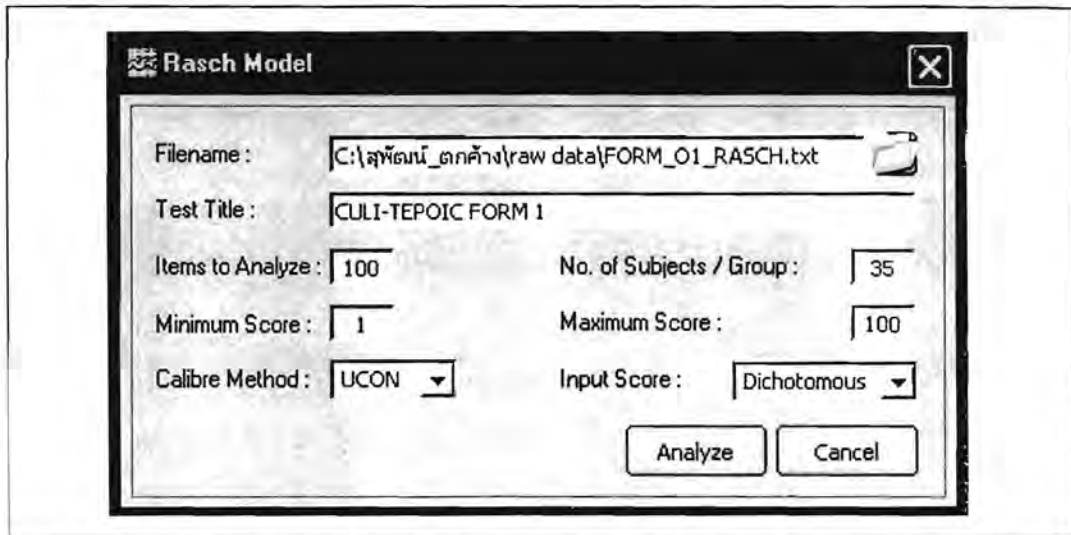
2. การใช้โปรแกรมวิเคราะห์ข้อทดสอบรายข้อแบบราช (Rasch Model)

หน้าจอของโปรแกรมวิเคราะห์ข้อทดสอบแบบราชมีรายละเอียดดังแสดงไว้ในภาพประกอบที่ 21 กล่าวคือจะประกอบด้วยปุ่มเปิดไฟล์เพื่อการนำเข้าข้อมูล ฟิลด์แสดงรายละเอียดข้อมูลที่จะทำการวิเคราะห์ และปุ่ม “Analyze” ซึ่งเป็นปุ่มที่สั่งให้โปรแกรมเริ่มทำการวิเคราะห์ข้อทดสอบปุ่มนี้จะไม่ทำงานจนกว่าจะมีการนำเข้าข้อมูลที่มีการจัดรูปแบบของข้อมูลที่ต้องการ



ภาพประกอบที่ 21: หน้าจอนำเข้าข้อมูลที่จะทำการวิเคราะห์ข้อทดสอบ

ในกรณีที่ข้อมูลนำเข้ามีรูปแบบที่ต้องการ โปรแกรมจะแสดงรายละเอียดของไฟล์นั้น และปุ่ม “Analyze” จะมีสีเข้มขึ้นกว่าเดิมซึ่งแสดงว่าโปรแกรมพร้อมที่จะสามารถใช้งานได้ ดังแสดงไว้ในภาพประกอบที่ 22



ภาพประกอบที่ 22: หน้าจอ เมื่อนำเข้าข้อมูลที่มีรูปแบบที่ถูกต้อง
 หนึ่ง ที่ฟิลด์ชื่อ “Test Title” ผู้ใช้สามารถใส่ชื่อของแบบทดสอบที่นำมาวิเคราะห์ได้ และ
 ชื่อนี้จะไปปรากฏอยู่ที่ไฟล์ของผลการวิเคราะห์ (output file) เมื่อใส่ข้อมูลครบถ้วนแล้วให้คลิกที่ปุ่ม
 “Analyze” เพื่อทำการวิเคราะห์ข้อทดสอบต่อไป แล้วโปรแกรมจะรายงานผลการวิเคราะห์ดังแสดงไว้ใน
 ภาพประกอบที่ 23



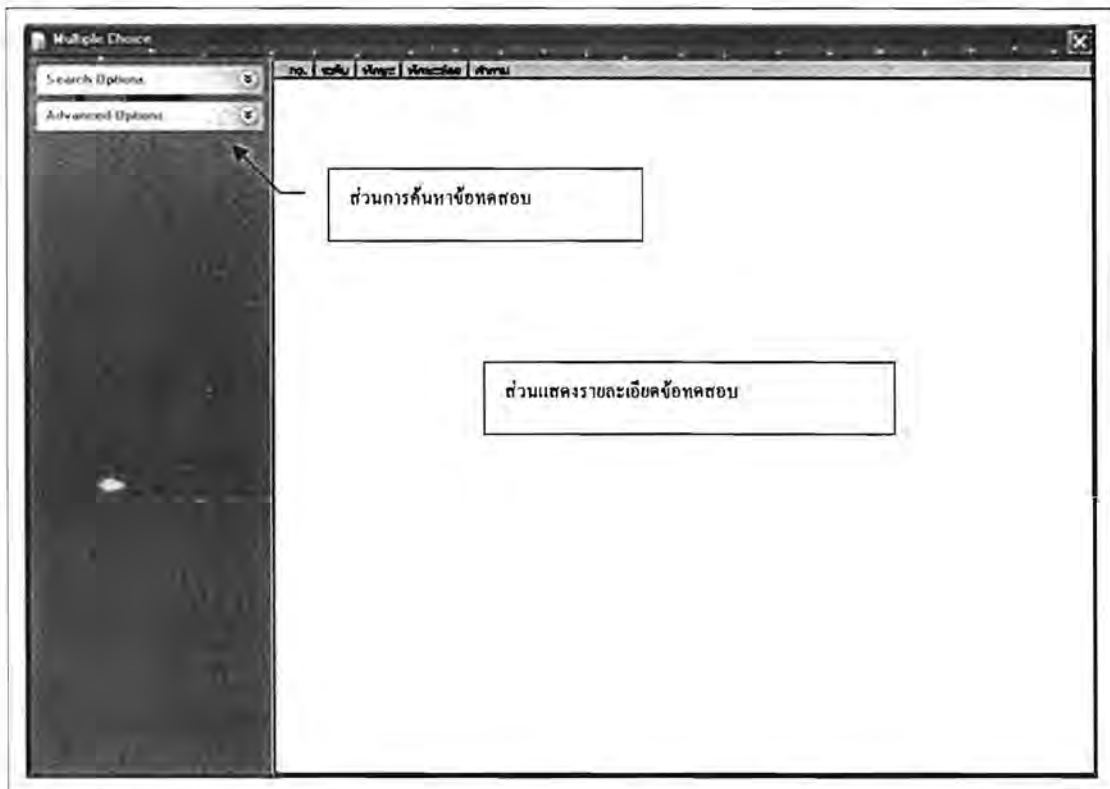
ภาพประกอบที่ 23: หน้าจอแสดงผลรายงานผลการวิเคราะห์ข้อทดสอบ

3. การใช้โปรแกรมธนาคารข้อทดสอบ (Item Bank)

การทดสอบในส่วนที่เกี่ยวข้องกับ โปรแกรมธนาคารข้อทดสอบผู้วิจัยแยกการทดสอบตามเมนูย่อยของเมนูหลักชื่อ “Item Bank” ดังนี้

ก. สร้าง/ปรับปรุงข้อทดสอบแบบเลือกตอบ

ในการทดสอบระบบผู้วิจัยได้จัดเตรียมข้อมูลของข้อทดสอบวิชาภาษาอังกฤษที่จะนำเข้ามาเก็บไว้ในธนาคารข้อทดสอบ ข้อทดสอบเหล่านี้ได้ทำการวิเคราะห์แล้ว โดยวิธีประเพณีนิยมจากโปรแกรมที่ 1 จึงมีค่าความยาก และค่าอำนาจจำแนก รวมทั้งรหัสอื่นๆที่จำเป็นต้องใช้ เมื่อผู้วิจัยได้เลือกเมนูย่อยชื่อ “สร้าง/ปรับปรุงข้อทดสอบแบบเลือกตอบ” จะปรากฏหน้าจอ ดังแสดงในภาพประกอบที่ 24



ภาพประกอบที่ 24: หน้าจอ เมื่อคลิกเมนูย่อยชื่อ “สร้าง/ปรับปรุงข้อทดสอบแบบเลือกตอบ”

ที่หน้าจอนี้ เมื่อคลิกเมาส์ที่ด้านขวา (right click) ที่บริเวณพื้นที่ส่วนแสดงรายละเอียดข้อทดสอบจะมีเมนูย่อยเกิดขึ้น และเมื่อเลือกที่เมนูย่อยชื่อ “New” โปรแกรมจะแสดงหน้าจอให้ผู้ใช้ป้อนรายละเอียดของข้อทดสอบ ดังแสดงไว้ในภาพประกอบที่ 25

ภาพประกอบที่ 25: หน้าจอแสดงการสร้างข้อทดสอบแบบเลือกตอบ

ผู้วิจัยได้ทดลองสร้างข้อทดสอบตามข้อมูลที่ได้จัดเตรียมไว้ เช่น ระบุระดับของข้อทดสอบว่าผู้สอบควรมีการศึกษาระดับใด ทักษะที่ต้องการทดสอบ (เช่น การอ่าน การฟัง และการเขียน เป็นต้น) และทักษะย่อยของข้อทดสอบ(เช่น อ่านจับใจความสำคัญ อ่านหารายละเอียด การอนุมาน และการเดาความหมายของคำศัพท์ เป็นต้น) ค่าความยาก ค่าอำนาจจำแนก ข้อคำถาม และตัวเลือก เป็นต้น การกรอกรายละเอียดในส่วนนี้สามารถทำได้ทั้งภาษาไทยและภาษาอังกฤษ และส่วนที่เป็นเรื่องเกี่ยวกับระดับ ทักษะและทักษะย่อย ผู้ใช้จะต้องเตรียมรายละเอียดส่วนนี้ไว้ก่อนในขั้นตอนการเตรียมรายละเอียดของระบบฐานข้อมูลของธนาคารข้อทดสอบแล้วมาทำการเลือกรายละเอียดเหล่านี้ในภายหลัง เมื่อกรอกรายละเอียดเสร็จแล้วจะมีลักษณะดังแสดงไว้ในภาพประกอบที่ 26 ต่อไปนี้

ภาพประกอบที่ 26: ตัวอย่างข้อมูลการสร้างข้อทดสอบแบบเลือกตอบ

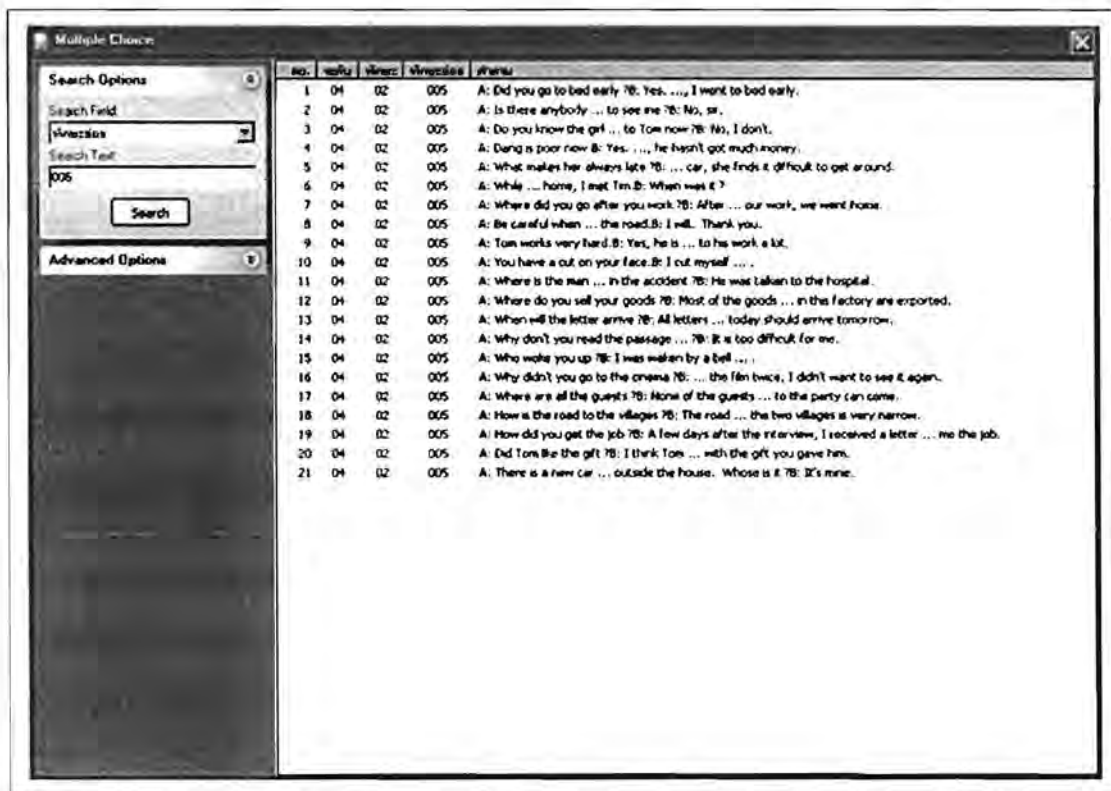
เมื่อต้องการบันทึกข้อมูลไว้ในธนาคารข้อทดสอบ ให้คลิกที่ปุ่ม “Save” และเมื่อโปรแกรมบันทึกข้อมูลเสร็จแล้วจะมีข้อความดังในภาพประกอบที่ 27 เกิดขึ้นเพื่อแจ้งให้ผู้ใช้ทราบว่า โปรแกรมได้ทำการบันทึกข้อทดสอบที่สร้างขึ้นไว้ในฐานข้อมูลเรียบร้อยแล้วดังนี้



ภาพประกอบที่ 27: แสดงข้อความ เมื่อการสร้างข้อทดสอบเป็นผลสำเร็จ

ข. การค้นหาข้อทดสอบที่ต้องการ

การค้นหาข้อทดสอบที่เก็บอยู่ในฐานข้อมูล สามารถทำได้โดยการกำหนดเงื่อนไขในการค้นหาที่อยู่ในส่วนการค้นหาข้อทดสอบที่แสดงไว้ในภาพประกอบที่ 28 โดยผู้ใช้โปรแกรมสามารถคลิกที่แถบหน้าต่างชื่อ “Search Options” ในส่วนการสืบค้นข้อทดสอบ แถบชื่อ “Search Options” จะเลื่อนลงมาให้เห็นส่วนการป้อนเงื่อนไขการค้นหา เมื่อผู้วิจัยได้ทดสอบการค้นหาข้อทดสอบด้วยการทดลองกำหนดเงื่อนไขการค้นหาด้วยการพิมพ์คำว่า “ทักษะย่อย” ลงในฟิลด์(field)ชื่อ “Search Field” และป้อนค่า “005” ลงในฟิลด์ “Search Text” เสร็จแล้วคลิกปุ่ม “Search” เพื่อให้โปรแกรมค้นหาข้อทดสอบที่สอดคล้องกับเงื่อนไขที่ระบุแล้วจะได้รายละเอียดดังแสดงไว้ในภาพประกอบที่ 28 ต่อไปนี้

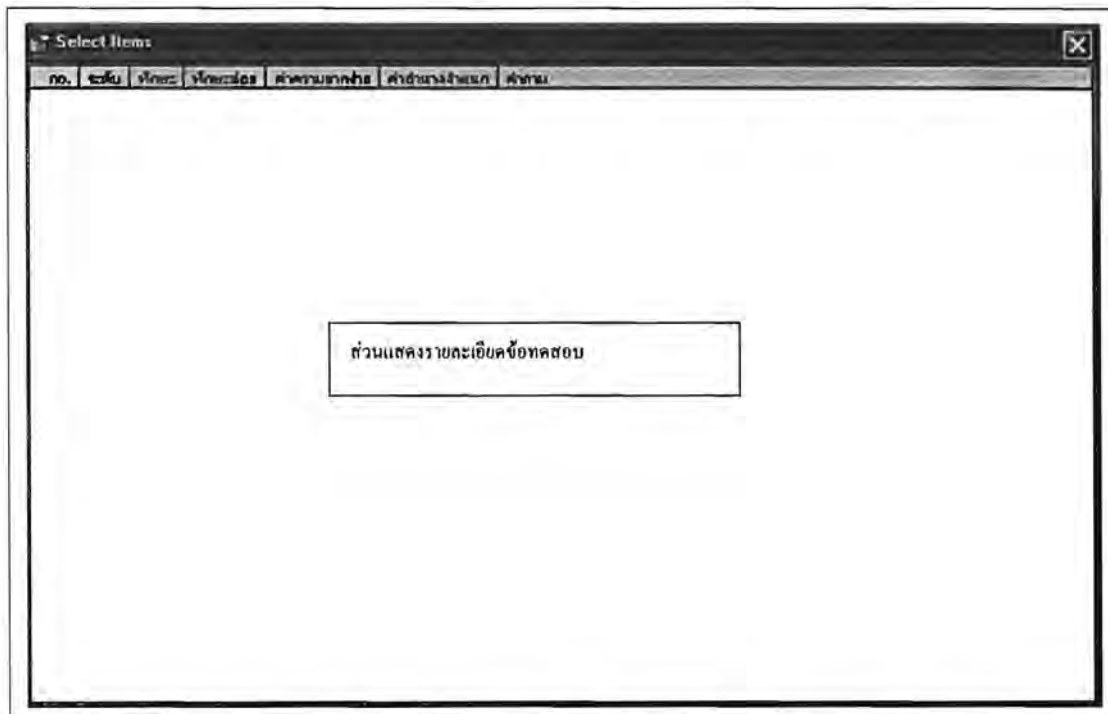


ภาพประกอบที่ 28: หน้าจอแสดงวิธีและผลการค้นหาข้อทดสอบ

ค. เมนูย่อย เลือกข้อทดสอบ/แบบทดสอบ

การเลือกข้อทดสอบที่จัดเก็บไว้ในฐานข้อมูลมาใช้งานมีวิธีการคล้ายคลึงกับการค้นหาข้อทดสอบในส่วนของการสร้าง/ปรับปรุงข้อทดสอบแบบเลือกตอบ แต่การค้นหาในส่วนของการเลือกข้อ

ทดสอบ/แบบทดสอบนั้น จุดประสงค์เพื่อนำข้อทดสอบที่อยู่ในฐานข้อมูลมาจัดพิมพ์เป็นแบบ ทดสอบ จึงสามารถกำหนดเงื่อนไขการค้นหาได้มากกว่าการค้นหาแบบแรก การทดสอบ โปรแกรมส่วนนี้เริ่มจากการคลิก(click)เลือกเมนูย่อยชื่อ “เลือกข้อทดสอบมาใช้งาน” จากเมนูหลักชื่อ “Item bank” ดังแสดงไว้ในภาพประกอบที่ 29



ภาพประกอบที่ 29: หน้าจอ เมื่อเลือกเมนูย่อยชื่อ “เลือกข้อทดสอบ/แบบทดสอบ”

เมื่อคลิกเมาส์ด้านขวา (right click) ที่หน้าจอนี้ บริเวณพื้นที่ส่วนแสดงรายละเอียดข้อทดสอบจะเกิดมีเมนูย่อยขึ้น และเมื่อเลือกที่เมนูย่อยชื่อ “Search” โปรแกรมจะแสดงหน้าจอให้ผู้ใช้ป้อนเงื่อนไขในการค้นหาข้อทดสอบ ดังตัวอย่างในภาพประกอบที่ 30 ต่อไปนี้

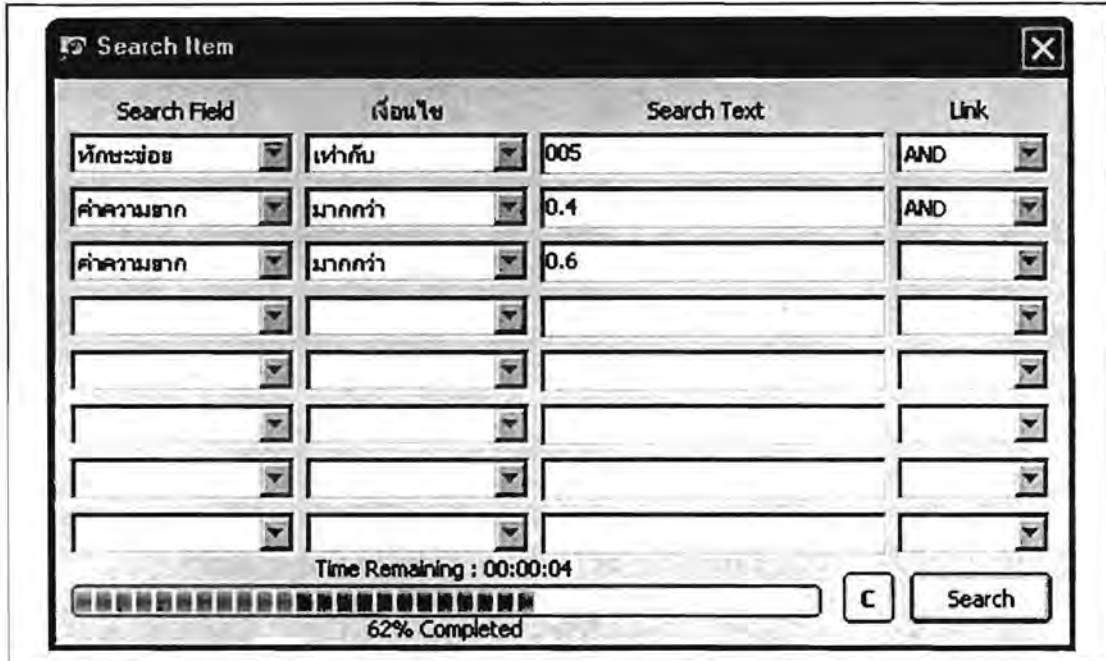
Search Field	เงื่อนไข	Search Text	Link
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

C Search

ภาพประกอบที่ 30: หน้าจอให้ผู้ใช้กำหนดเงื่อนไขในการค้นหาข้อทดสอบ

การใช้โปรแกรมในส่วนนี้ ทำได้โดยการคลิกที่ปุ่มสามเหลี่ยมสีดำในแต่ละฟิลด์เพื่อกำหนดเงื่อนไขที่ต้องการ เช่น ทักษะย่อยเท่ากับ 005 และ (AND) ค่าความยากเท่ากับ 0.4 และ (AND) ค่าความยากมากกว่า 0.6 เป็นต้น อนึ่ง เงื่อนไขเหล่านี้จะระบุละเอียดมากน้อยเพียงใดแล้วแต่ความต้องการของผู้ใช้ว่าต้องการได้ข้อทดสอบหรือแบบทดสอบที่มีรายละเอียดเฉพาะมากน้อยเพียงใด

เมื่อระบุเงื่อนไขในการค้นหาเรียบร้อยแล้วและคลิกที่ปุ่ม “Search” โปรแกรมจะทำการค้นหาข้อทดสอบที่สอดคล้องกับเงื่อนไข ผู้ใช้สามารถสังเกตเห็นได้จากแถบสีเขียวด้านข้างของหน้าต่าง ดังแสดงไว้ในภาพประกอบที่ 31 ต่อไปนี้



ภาพประกอบที่ 31: หน้าจอ แสดงสถานการณ์ทำงานกรณีพบข้อทดสอบตามเงื่อนไขที่ระบุ

ในกรณีที่เงื่อนไขที่ผู้ใช้โปรแกรมกำหนดขึ้น โปรแกรมไม่สามารถค้นหาข้อทดสอบหรือแบบทดสอบได้ (เช่น ไม่มีข้อทดสอบหรือแบบทดสอบตามเงื่อนไข) โปรแกรมจะมีข้อความบอกผู้ใช้ให้ทราบว่าไม่มีข้อมูลที่กำหนดตามเงื่อนไข ดังแสดงไว้ในภาพประกอบที่ 32 ต่อไปนี้



ภาพประกอบที่ 32: แสดงข้อความ เมื่อไม่พบข้อทดสอบตามเงื่อนไขที่ระบุ

4. การใช้โปรแกรมการทดสอบแบบปรับเปลี่ยน (CCAT)

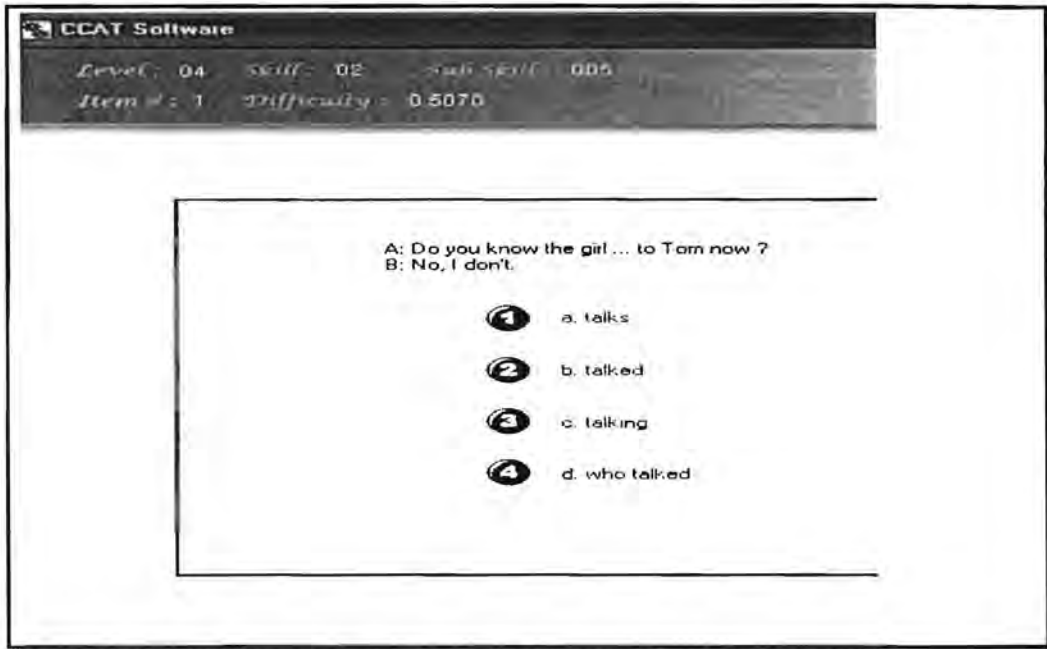
การใช้โปรแกรมทดสอบแบบปรับเปลี่ยน ผู้ใช้จะต้องเปิด โปรแกรมที่ชื่อว่า CCAT.exe ขึ้นด้วยการคลิกสองครั้ง(double click) ที่ Icon ชื่อ CCAT (Computerized Content-based Adaptive Testing) แล้วโปรแกรมจะแสดงหน้าต่างตามภาพประกอบที่ 33 ขึ้น ให้กรอกข้อความลงในแต่ละฟิลด์แล้วคลิกที่ปุ่ม “Enter”

ภาพประกอบที่ 33: หน้าจอให้กรอกข้อมูลของผู้ประสงค์จะทำการทดสอบแบบปรับเปลี่ยน

ขั้นต่อไปเป็นการระบุจำนวนปริเขต (domain) หรือแบบทดสอบย่อยเฉพาะเนื้อหา และ รายละเอียดของแบบทดสอบที่ต้องการทดสอบ โดยผู้ใช้โปรแกรมสามารถกำหนดรายละเอียดเกี่ยวกับระดับของแบบทดสอบ ทักษะ และทักษะย่อยได้จากการคลิกที่รูปสามเหลี่ยมสีดำ เสร็จแล้วคลิกที่ปุ่ม “Go” เพื่อให้โปรแกรมทำงานต่อไป ดังรายละเอียดในภาพประกอบที่ 34

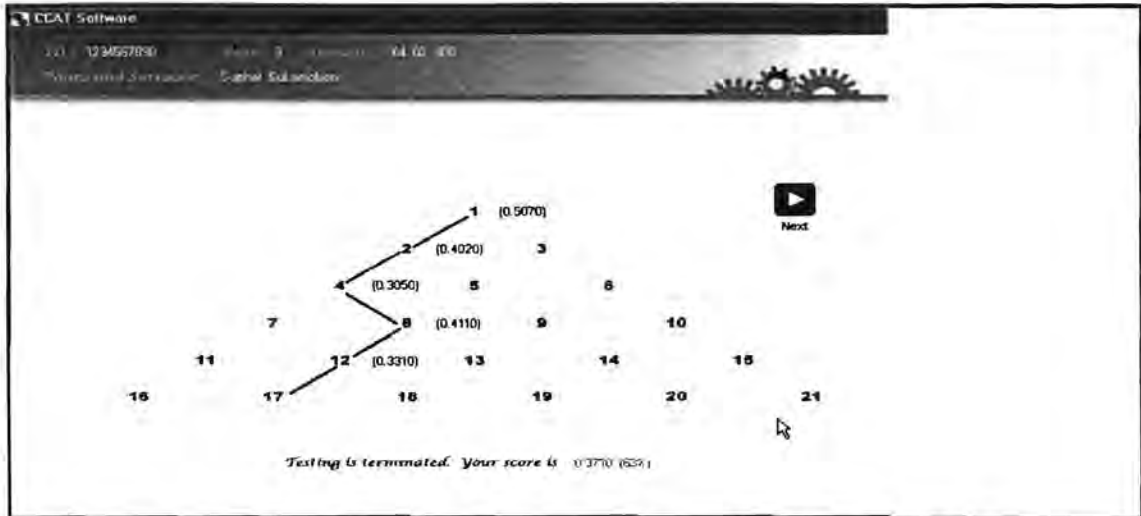
ภาพประกอบที่ 34: แสดงการเลือกปริเขตและรายละเอียดของแบบทดสอบแบบปรับเปลี่ยน

หลังจากนั้น โปรแกรมจะทำการสุ่มเลือกข้อทดสอบที่มีลักษณะสอดคล้องกับเงื่อนไขที่ ระบุไว้ในขั้นตอนข้างต้นมาให้ผู้สอบตอบทีละข้อจากธนาคารข้อทดสอบ ผู้สอบตอบข้อทดสอบโดยการคลิกเลือกคำตอบที่ต้องการและคลิกปุ่ม "Next" เพื่อยืนยันคำตอบ แล้วโปรแกรมจะสุ่มเลือกข้อทดสอบข้อต่อไปมาให้ผู้สอบตอบ ลักษณะของข้อทดสอบข้อต่อไปจะยากหรือง่ายขึ้นอยู่กับผลการตอบของผู้สอบข้อทดสอบนี้ว่าตอบถูกหรือผิด ดังแสดงไว้ในภาพประกอบที่ 35 ต่อไปนี้



ภาพประกอบที่ 35: ลักษณะของข้อทดสอบแบบปรับเปลี่ยนตามเงื่อนไขที่ระบุ

เมื่อสิ้นสุดการทดสอบแต่ละปริเซต โปรแกรมจะแสดงวิธีการทดสอบ (testing path) ของผู้สอบว่ามีลักษณะอย่างไร ดังแสดงไว้ในภาพประกอบที่ 36



ภาพประกอบที่ 36: วิธีการทดสอบของผู้สอบแต่ละคนในแต่ละปริเขต

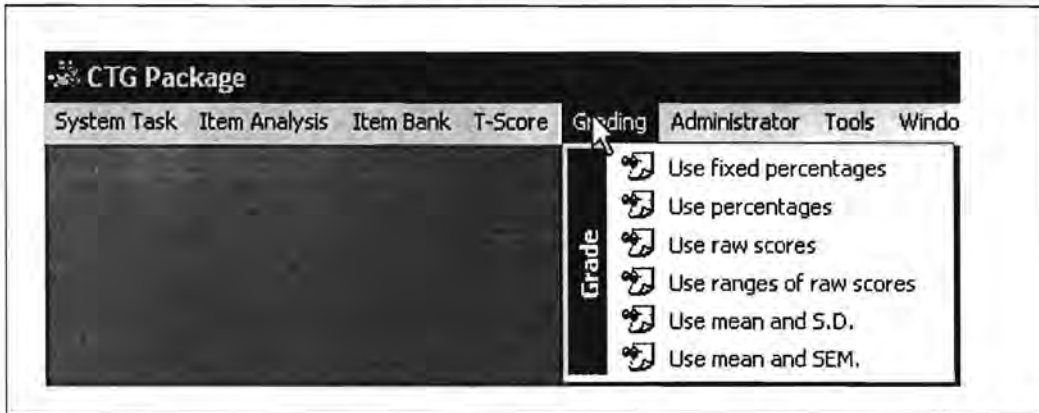
เมื่อการทดสอบทุกปริเขตสิ้นสุดแล้ว โปรแกรมจะแสดงผลการทดสอบให้แก่ผู้สอบทราบ ผู้สอบสามารถจะพิมพ์ผลการทดสอบนี้ได้ และผลการทดสอบนี้ผู้บริหารระบบการทดสอบสามารถที่จะตรวจสอบได้ในภายหลังเพื่อประโยชน์ในการให้คำแนะนำเกี่ยวกับการเรียนแก่ผู้สอบได้ รายละเอียดของผลการทดสอบและเกณฑ์แสดงไว้ในภาพประกอบที่ 37

Testing Report			
ID: 1234567890	Year: 3		
Name and Surname: Suphat Sukamolson			
Testing domain (s) and score (s)			
1.) 04 02 005 0.3710	2.) 04 02 006 0.0000	3.) 00 00 000 0.0000	
4.) 00 00 000 0.0000	5.) 00 00 000 0.0000	6.) 00 00 000 0.0000	
7.) 00 00 000 0.0000	8.) 00 00 000 0.0000	9.) 00 00 000 0.0000	
10.) 00 00 000 0.0000	11.) 00 00 000 0.0000	12.) 00 00 000 0.0000	
13.) 00 00 000 0.0000	14.) 00 00 000 0.0000	15.) 00 00 000 0.0000	
16.) 00 00 000 0.0000	17.) 00 00 000 0.0000	18.) 00 00 000 0.0000	
19.) 00 00 000 0.0000	20.) 00 00 000 0.0000	21.) 00 00 000 0.0000	
22.) 00 00 000 0.0000	23.) 00 00 000 0.0000	24.) 00 00 000 0.0000	
25.) 00 00 000 0.0000			
Average Score =0.1855 (81%)			
Evaluation criteria			
0% - 20%	Very poor		
21% - 40%	Poor		
41% - 60%	Moderate		
61% - 80%	Good		
81% - 100%	Excellent		

ภาพประกอบที่ 37: คะแนนผลการทดสอบแบบทดสอบแบบปรับเปลี่ยนทุกปริเขตรวมกัน

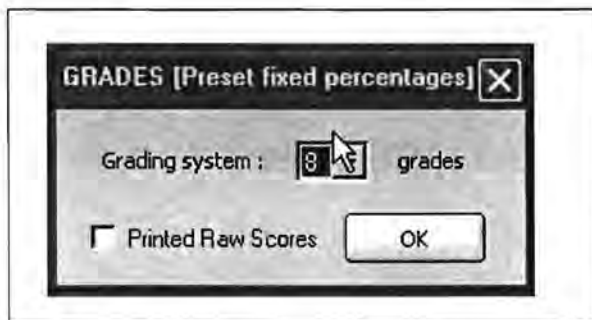
5. การใช้โปรแกรมกำหนดระดับคะแนน (Grading and T-Score)

การใช้โปรแกรมกำหนดระดับคะแนนทำได้โดยการคลิกที่คำสั่งชื่อ “Grading” ที่แถบคำสั่ง (menu bar) แล้วจะมีรายการให้เลือก 6 อย่าง และหากว่าคลิกที่คำสั่ง “T-Score” ก็จะพบวิธีการกำหนดระดับคะแนนโดยใช้คะแนนมาตรฐานที่ รวมแล้วเป็นทั้งหมด 7 วิธี ดังภาพประกอบที่ 38



ภาพประกอบที่ 38: รายการวิธีการกำหนดระดับคะแนนแบบต่างๆ

เมื่อผู้ใช้โปรแกรมเลือกวิธีการกำหนดระดับคะแนนแต่ละวิธีแล้ว โปรแกรมจะถามว่าต้องการกำหนดระดับคะแนนเป็นแบบ 5 ระดับ หรือ 8 ระดับ และต้องการให้โปรแกรมพิมพ์คะแนนดิบด้วยหรือไม่ นอกเหนือจากระดับคะแนนแล้ว รายละเอียดดังแสดงในภาพประกอบที่ 39



ภาพประกอบที่ 39: รายละเอียดการกำหนดระดับคะแนนแต่ละวิธี

ภายหลังจากการคลิกที่ปุ่ม “OK” แล้ว โปรแกรมย่อยเพื่อกำหนดระดับคะแนนแต่ละวิธีมีรายละเอียดแตกต่างกัน และผลลัพธ์ที่ได้อาจแตกต่างกันหากว่าเกณฑ์ในการกำหนดระดับคะแนนแตกต่างกัน แต่รูปแบบของการรายงานผลการกำหนดระดับคะแนนแต่ละวิธีจะเหมือนกัน ดังแสดงไว้ในภาพประกอบที่ 40 และ 41 ต่อไปนี้

Grading Tech

This is a program for assigning grades into 2 systems :

a. 5 GRADES: A, B, C, D and F

b. 8 GRADES: A, B+, B, C+, C, D+, D and F

It was originally written for DOS by

Dr. Suphat Sukamolson, Ph.D. *

(Educational Measurement and Evaluation),

and updated for Windows by

Mr. Komson Eiamjamrus, M.Sc.

LANGUAGE INSTITUTE

CHULALONGKORN UNIVERSITY, 2004

ALL RIGHTS RESERVED

[* Sukamolson, S. *Computerized Item Analyses and Grading*: Wittayaphat, 2542.]

ภาพประกอบที่ 40: หน้าแรกของผลการกำหนดระดับคะแนนแต่ละวิธี

Like These Files Properties

COURSE TITLE _____

No.	Name and Last Name	Score	Grade
112	033	36	F
113	071	33	F
114	046	31	F
115	050	29	F
116	051	25	F

A = 7	D+ = 18
B+ = 14	D = 12
B = 15	F = 5
C+ = 29	TOTAL = 116
C = 16	

Instructor's Signature: _____

Date: ____/____/____

ภาพประกอบที่ 41: หน้าสุดท้ายของผลการกำหนดระดับคะแนนแต่ละวิธี

บทที่ 4

ผลการวิจัย

ผลของการวิจัยตามวัตถุประสงค์ทั้ง 2 มีรายละเอียดดังต่อไปนี้

ส่วนที่ 1 การสร้างและพัฒนาแบบทดสอบ CULI-TEPOIC

1. ค่าสถิติพื้นฐานของแบบทดสอบแต่ละชุด

ตารางที่ 11: ค่าสถิติพื้นฐานของแบบทดสอบแต่ละชุด

ค่าสถิติ	CULI-TEPOIC	CULI-TEPOIC	CULI-TEPOIC
	ชุดที่ 1	ชุดที่ 2	ชุดที่ 3
จำนวนข้อทดสอบ	100	100	100
จำนวนผู้สอบ	116	124	120
ค่าคะแนนเฉลี่ย	57.87	55.32	53.69
ค่าเบี่ยงเบนมาตรฐาน	13.30	16.15	16.20
ค่าความคลาดเคลื่อนมาตรฐานค่าเฉลี่ย	1.235	1.451	1.479
คะแนนสูงสุด	90	92	93
คะแนนต่ำสุด	25	25	22
พิสัยของคะแนน	65	67	71
ค่าเบี่ยงเบนควอร์ไทล์	9.75	12.50	12.25
ค่ามัธยฐาน	57	55	53
ค่าฐานนิยม	57**	47**	47**
ค่าความเบ้	-0.02	0.09	0.19
ค่าความโด่ง	-0.44	-0.88	-0.69
ค่าดัชนีการกระจายของคะแนน	22.98%	29.19%	30.17%

** ค่าฐานนิยมเป็นค่าโดยประมาณหากว่าการกระจายไม่เป็น โค้งปกติ

จากตารางที่ 11 แสดงให้เห็นว่า โดยทั่วไปแล้วค่าสถิติพื้นฐานต่างๆของแบบทดสอบ CULI-TEPOIC แต่ละชุดมีลักษณะใกล้เคียงกันมาก เช่น คะแนนสูงสุด คะแนนต่ำสุด ค่าความเบ้ ค่าความโด่งของ

แต่ละชุดใกล้เคียงกัน และค่าคะแนนเฉลี่ย ค่าเบี่ยงเบนมาตรฐาน และค่าความคลาดเคลื่อนมาตรฐานค่าเฉลี่ย (standard error of mean) แตกต่างกันไม่มากนัก ค่าสถิติพื้นฐานเหล่านี้แสดงให้เห็นว่า โดยภาพรวมแล้ว แบบทดสอบทั้ง 3 ชุดมีลักษณะคล้ายคลึงกันมาก และที่สำคัญมากก็คือ ค่าความเบ้ที่ใกล้เคียงกับ 0.0 และค่าความโด่งที่ติดลบและต่างจาก 1.0 ไม่มากนัก แสดงให้เห็นว่า การกระจายของคะแนนสอบมีแนวโน้มเป็นโค้งปกติ ลักษณะดังกล่าวนี้ใช้เป็นเกณฑ์ในการเลือกใช้ค่าสัมประสิทธิ์สหสัมพันธ์แบบพอยท์ไบเซเรียล (Point-Beserial Correlation Coefficient: r_{pb}) เป็นค่าความตรงของข้อทดสอบ (Test Item Validity Index) ในภายหลัง

2. ค่าสถิติโดยสรุปของแบบทดสอบทั้ง 3 ชุด

ตารางที่ 12: ค่าสถิติโดยสรุปของแบบทดสอบแต่ละชุด

ค่าสถิติ	CULI-TEPOIC	CULI-TEPOIC	CULI-TEPOIC
	ชุดที่ 1	ชุดที่ 2	ชุดที่ 3
จำนวนข้อทดสอบทั้งหมด	100	100	100
\bar{X} ของคะแนนสอบ	57.871	55.323	53.692
\bar{X} ของค่าความยาก	0.579	0.553	0.537
\bar{X} ของค่าความยากแบบ Δ	12.029	12.376	12.579
\bar{X} ของค่าอำนาจจำแนก	0.334	0.405	0.403
\bar{X} ของค่าความตรงรายข้อแบบ r_{bis}	0.383	0.459	0.441
\bar{X} ของค่าความตรงรายข้อแบบ r_{pb}	0.292	0.353	0.343
ค่าความเที่ยงแบบ KR_{20}	0.893	0.930	0.926
ค่าความเที่ยงแบบ KR_{21}	0.871	0.914	0.914
ค่าความเที่ยงแบบ Alpha	0.893	0.930	0.926
ค่าความเที่ยงแบบ Split-half	0.865	0.925	0.925
จำนวนข้อทดสอบที่ควรปรับปรุง	15	12	15

ตารางที่ 12 แสดงให้เห็นอย่างชัดเจนว่าแบบทดสอบ CULI-TEPOIC แต่ละชุดมีลักษณะทั่วไปโดยสรุปใกล้เคียงกันมาก เช่น มีค่าความยากโดยเฉลี่ยอยู่ในระดับปานกลางเหมือนกัน ($p = 0.553 - 0.579$) มีค่าอำนาจจำแนกโดยเฉลี่ยอยู่ในระดับที่ดี (DIIA, 2003) และใกล้เคียงกันโดยเฉพาะชุดที่ 2 และ 3 ($d = 0.334 - 0.405$) มีค่าความตรงรายข้อแบบ r_{pb} โดยเฉลี่ยอยู่ในระดับที่ดี (DIIA, 2003) และใกล้เคียงกันโดยเฉพาะชุดที่

2 และ 3 ($r_{pb} = 0.292 - 0.353$) และแบบทดสอบ CULI-TEPOIC ชุดที่ 1 มีค่าความเที่ยงอยู่ในระดับดีมาก และชุดที่ 2 และ 3 มีค่าความเที่ยงอยู่ในระดับดีขอดีเยี่ยม(DIA, 2003) แต่แบบทดสอบทั้ง 3 ชุดต่างก็มีค่าดังกล่าวแตกต่างกันไม่มากนัก ($KR_{20} = 0.893 - 0.930$)

อนึ่ง จากการวิเคราะห์ข้อทดสอบรายข้อพบว่า แบบทดสอบ CULI-TEPOIC ชุดที่ 1, 2 และ 3 มีข้อทดสอบที่ควรได้รับการปรับปรุงแก้ไขให้มีคุณภาพดียิ่งขึ้นจำนวน 15, 12 และ 15 ข้อตามลำดับ ซึ่งอาจกล่าวได้ว่าเป็นจำนวนที่ไม่มากนักเมื่อเทียบกับคุณภาพของข้อทดสอบแต่ละชุดโดยรวม (รายละเอียดของผลการวิเคราะห์ข้อทดสอบรายข้อสามารถพบได้ในภาคผนวก ก.)

3. ค่าความตรงร่วมสมัย (concurrent validity) ของแบบทดสอบทั้ง 3 ชุด และสมการพยากรณ์

ตารางที่ 13: ค่าความตรงร่วมสมัย (concurrent validity) ของแบบทดสอบทั้ง 3 ชุด

Form	Beta	a (constant)	b	error	n of testees
CULI-TEPOIC 1	0.976	-88.791	12.001	34.990	24
CULI-TEPOIC 2	0.943	-14.464	10.766	52.659	22
CULI-TEPOIC 3	0.904	16.379	10.610	71.537	19

จากตารางที่ 13 แสดงให้เห็นว่าค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนผู้สอบแบบทดสอบ CULI-TEPOIC ชุดที่ 1, 2 และ 3 กับคะแนนสอบแบบทดสอบ TOEIC คือ r_{xy} (หรือ Beta) = 0.976, 0.943 และ 0.904 ตามลำดับ ซึ่งมีค่าใกล้เคียงกันและอยู่ในระดับสูงและสูงมาก และจากค่าสถิติต่างในตาราง 13 สามารถเขียนเป็นสมการเพื่อพยากรณ์คะแนนสอบ TOEIC จากการทดสอบแบบทดสอบ CULI-TEPOIC ชุดที่ 1, 2 และ 3 ได้ดังนี้

CULI-TEPOIC ชุดที่ 1

$$y' = 12.001 \times \text{คะแนน CULI-TEPOIC 1} + (-88.791)$$

CULI-TEPOIC ชุดที่ 2

$$y' = 10.766 \times \text{คะแนน CULI-TEPOIC 2} + (-14.464)$$

CULI-TEPOIC ชุดที่ 3

$$y' = 10.610 \times \text{คะแนน CULI-TEPOIC 3} + 16.376$$

(รายละเอียดของตารางเปรียบเทียบคะแนนของแบบทดสอบทั้ง 2 ชนิดสามารถพบได้ในภาคผนวก

ง.)

ดังนั้น จึงสามารถกล่าวสรุปได้ว่า แบบทดสอบ CULI-TEPOIC ทั้ง 3 ชุดเป็นแบบทดสอบวัดสมรรถภาพทั่วไปทางภาษาอังกฤษเพื่อวิชาชีพและการสื่อสารนานาชาติที่มีคุณภาพดี ทั้งในด้านคุณภาพของข้อทดสอบรายข้อ ค่าความเที่ยง และค่าความตรงร่วมสมัยกับแบบทดสอบที่มีชื่อเสียงและเมื่อปรับปรุงแก้ไขข้อทดสอบชุดละประมาณ 12-15 ข้อที่ยังไม่ดีได้แล้ว จะทำให้แบบทดสอบทั้ง 3 ชุดเป็นแบบทดสอบที่มีคุณภาพดีมาก

3. ค่าความยากของข้อทดสอบ (β) ทั้ง 3 ชุด เมื่อวิเคราะห์ข้อทดสอบโดยใช้ Rasch Model

เพื่อประโยชน์ในการจัดเรียงลำดับข้อทดสอบแต่ละข้อตามลำดับค่าความยากเมื่อนำข้อทดสอบไปใช้จริงในภายหลัง ผู้วิจัยจึงทำการวิเคราะห์ข้อทดสอบของแบบทดสอบทั้ง 3 ชุดอีกครั้งหนึ่งโดยใช้ Rasch Model (สุพัตน์ สุขมลสันต์, 2546) ทั้งนี้เพราะค่าที่ได้จากการวิเคราะห์ดังกล่าวนี้เป็นค่าที่ไม่เปลี่ยนแปลงตามลักษณะของผู้สอบ (Lord, 1980) และผลของการวิเคราะห์ที่ได้แสดงไว้แล้วในภาคผนวก ค. (เนื่องจากตารางมีขนาดใหญ่และมีรายละเอียดมาก)

2. ส่วนที่ 2 การสร้างและพัฒนาโปรแกรมคอมพิวเตอร์ที่ใช้ในการพัฒนาแบบทดสอบและการทดสอบ

1. การเปรียบเทียบผลการทำงานของโปรแกรมวิเคราะห์ข้อทดสอบแบบประเพณีนิยมที่พัฒนาขึ้น ซึ่งเรียกชื่อว่าโปรแกรม Classical กับโปรแกรม CTIA 7.0 (Classical Test Item Analysis)

เพื่อทดสอบการทำงานของโปรแกรมวิเคราะห์ข้อทดสอบแบบประเพณีนิยมที่พัฒนาขึ้น (โปรแกรม Classical) ว่าถูกต้องหรือไม่ ผู้วิจัยได้นำข้อมูลจริงจากการทดลองสอบแบบทดสอบ CULI-TEPOIC ชุดที่ 1 มาทดสอบใช้กับ โปรแกรมวิเคราะห์ข้อทดสอบที่พัฒนาขึ้น แล้วเปรียบเทียบที่ได้กับผลการวิเคราะห์จากโปรแกรม CTIA 7.0 ซึ่งผู้วิจัยได้พัฒนาขึ้นใช้ก่อนแล้ว ปรากฏผลดังนี้

ตารางที่ 14: การเปรียบเทียบหาค่าสถิติพื้นฐานเมื่อคำนวณโดยโปรแกรม Classical กับ โปรแกรม CTIA 7.0 จากข้อมูลชุดเดียวกัน

ค่าสถิติพื้นฐาน *	จาก โปรแกรม Classical	จาก โปรแกรม CTIA 7.0
จำนวนข้อทดสอบ	100	100
จำนวนผู้สอบ	116	116
ค่าคะแนนเฉลี่ย	57.87	57.87
ค่าเบี่ยงเบนมาตรฐาน	13.30	13.30
ค่าความคลาดเคลื่อนมาตรฐานค่าเฉลี่ย	1.235	1.235
คะแนนสูงสุด	90	90
คะแนนต่ำสุด	25	25
พิสัยของคะแนน	65	65
ค่าเบี่ยงเบนควอร์ไทล์	9.75	9.75
ค่ามัธยฐาน	57	57
ค่าฐานนิยม	57**	57**
ค่าความเบ้	-0.02	-0.02
ค่าความโด่ง	-0.44	-0.44

* $r_{xy} = 1.00$ ($p < 0.05$)

จากการเปรียบเทียบค่าสถิติพื้นฐานที่ได้จากการคำนวณโดยโปรแกรม Classical กับ โปรแกรม CTIA 7.0 จากข้อมูลชุดเดียวกันด้วยตาเปล่าในตารางที่ 14 ปรากฏว่าทุกค่าในมิติต่างๆ เช่น ค่าคะแนนเฉลี่ย ค่าเบี่ยงเบนมาตรฐาน ค่าฐานนิยม หรือค่าความเบ้ มีขนาดเท่ากันพอดี และเมื่อทำการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าสถิติพื้นฐานที่ได้จากโปรแกรมทั้งสองปรากฏว่า $r_{xy} = 1.00$ ($p < 0.05$) หมายความว่าทุกค่ามีขนาดเท่ากันพอดี ซึ่งแสดงว่าโปรแกรม Classical ทำการคำนวณได้ค่าสถิติพื้นฐานถูกต้องเช่นเดียวกับโปรแกรม CTIA 7.0

ตารางที่15: การเปรียบเทียบผลการวิเคราะห์ข้อทดสอบเป็นรายข้อ เมื่อคำนวณโดยโปรแกรม Classical กับโปรแกรม CTIA 7.0 จากข้อมูลชุดเดียวกัน

Item #	Choice #	จากโปรแกรม Classical			จากโปรแกรม CTIA 7.0		
		p *	d *	r_{pb} *	p	d	r_{pb}
1	*1	0.440	0.065	0.063	0.440	0.065	0.063
	2	0.112	-0.290	-0.319	0.112	-0.290	-0.319
	3	0.431	0.0258	0.152	0.431	0.0258	0.152
	4	0.017	-0.032	-0.049	0.017	-0.032	-0.049
	5	0.000	0.000	0.000	0.000	0.000	0.000
	ERROR	0.000	0.000	0.000	0.000	0.000	0.000
2	*1	0.440	0.548	0.400	0.440	0.548	0.400
	2	0.069	-0.065	-0.059	0.069	-0.065	-0.059
	3	.259	-0.226	-0.182	0.259	-0.226	-0.182
	4	0.224	-0.290	-0.279	0.224	-0.290	-0.279
	5	0.009	0.032	0.134	0.009	0.032	0.134
	ERROR	0.000	0.000	0.000	0.000	0.000	0.000

* $r_{xy} = 1.00$ ($p < 0.05$)

จากการเปรียบเทียบด้วยตาเปล่าผลการวิเคราะห์ข้อทดสอบเป็นรายข้อที่ได้จากการคำนวณโดยโปรแกรม Classical กับโปรแกรม CTIA 7.0 จากข้อมูลชุดเดียวกันในตารางที่ 15 ปรากฏว่าทุกค่าในมิติต่างๆ เช่น ค่าความยาก(p) ค่าอำนาจจำแนก(d) และค่าความตรงรายข้อ (r_{pb}) ของตัวเลือกต่างๆในแต่ละข้อมีขนาดเท่ากันพอดี และเมื่อทำการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าสถิติต่างๆที่ได้จากโปรแกรมทั้งสองปรากฏว่า $r_{xy} = 1.00$ ($p < 0.05$) หมายความว่าทุกค่ามีขนาดเท่ากันพอดี ซึ่งแสดงว่าโปรแกรม Classical ทำการคำนวณได้ค่าสถิติต่างๆในการวิเคราะห์ข้อทดสอบรายข้อแบบประเพณีนิยมได้ถูกต้องเช่นเดียวกับโปรแกรม CTIA 7.0

ตารางที่ 16: การเปรียบเทียบค่าสถิติโดยสรุปจากการวิเคราะห์ข้อทดสอบรายข้อเมื่อคำนวณโดยโปรแกรม Classical กับโปรแกรม CTIA 7.0 จากข้อมูลชุดเดียวกัน

ค่าสถิติ *	จากโปรแกรม Classical	จากโปรแกรม CTIA 7.0
จำนวนข้อทดสอบทั้งหมด	100	100
\bar{X} ของคะแนนสอบ	57.871	57.871
\bar{X} ของค่าความยาก	0.579	0.579
\bar{X} ของค่าความยากแบบ Δ	12.029	12.029
\bar{X} ของค่าอำนาจจำแนก	0.334	0.334
\bar{X} ของค่าความตรงรายข้อแบบ r_{bis}	0.383	0.383
\bar{X} ของค่าความตรงรายข้อแบบ r_{pb}	0.292	0.292
ค่าความเที่ยงแบบ KR_{20}	0.893	0.893
ค่าความเที่ยงแบบ KR_{21}	0.871	0.871
ค่าความเที่ยงแบบ Alpha	0.893	0.893
ค่าความเที่ยงแบบ Split-half	0.865	0.865
จำนวนข้อทดสอบที่ควรปรับปรุง	15	15

* $r_{xy} = 1.00$ ($p < 0.05$)

จากการเปรียบเทียบด้วยตาเปล่าผลการวิเคราะห์ข้อทดสอบเป็นรายข้อโดยสรุปที่ได้จากการคำนวณโดยโปรแกรม Classical กับโปรแกรม CTIA 7.0 จากข้อมูลชุดเดียวกันในตารางที่ 25 ปรากฏว่าทุกค่าในมิติต่างๆ เช่น ค่าความยากเฉลี่ย ค่าอำนาจจำแนกเฉลี่ย ค่าความตรงรายข้อเฉลี่ย และค่าความเที่ยงแบบต่างๆ มีขนาดเท่ากันพอดี และเมื่อทำการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าสถิติต่างๆที่ได้จากโปรแกรมทั้งสองปรากฏว่า $r_{xy} = 1.00$ ($p < 0.05$) หมายความว่าทุกค่ามีขนาดเท่ากันพอดี ซึ่งแสดงว่าโปรแกรม Classical ทำการคำนวณได้ค่าสถิติต่างๆโดยสรุปจากการวิเคราะห์ข้อทดสอบรายข้อแบบประเพณีนิยมได้ถูกต้องเช่นเดียวกับโปรแกรม CTIA 7.0

ดังนั้น จากผลการเปรียบเทียบค่าสถิติต่างๆที่ได้จากการใช้โปรแกรม Classical กับโปรแกรม CTIA 7.0 เพื่อการวิเคราะห์ข้อทดสอบรายข้อแบบประเพณีนิยมในตารางที่ 14-16 ข้างต้น ซึ่งได้ค่าเท่ากันพอดีทำให้สามารถสรุปได้ว่าโปรแกรม Classical ทำงานได้ถูกต้องเช่นเดียวกับโปรแกรม CTIA 7.0

2. การเปรียบเทียบผลการทำงานของโปรแกรมวิเคราะห์ข้อทดสอบแบบราชที่พัฒนาขึ้น ซึ่งเรียกชื่อว่า โปรแกรม Rasch กับโปรแกรม BICAL 7.0 (Binary Calibrator)

เพื่อทดสอบการทำงานของโปรแกรมวิเคราะห์ข้อทดสอบแบบราชที่พัฒนาขึ้น (โปรแกรม Rasch) ว่าถูกต้องหรือไม่ ผู้วิจัยได้นำข้อมูลจริงจากการทดลองสอบแบบทดสอบ CULI-TEPOIC ชุดที่ 1 มาทดสอบใช้กับโปรแกรมวิเคราะห์ข้อทดสอบที่พัฒนาขึ้น แล้วเปรียบผลที่ได้กับผลการวิเคราะห์จาก โปรแกรม BICAL 7.0 ที่ผู้วิจัยได้พัฒนาขึ้นใช้ก่อนแล้ว ปรากฏผลดังนี้

ตารางที่ 17: การเปรียบเทียบหาค่าสถิติพื้นฐานเมื่อคำนวณโดยโปรแกรม Rasch กับโปรแกรม BICAL 7.0 จากข้อมูลชุดเดียวกัน

ค่าสถิติพื้นฐาน *	จากโปรแกรม Rasch	จากโปรแกรม BICAL 7.0
จำนวนข้อทดสอบ	100	100
จำนวนผู้สอบ	232 (116*2)	232 (116*2)
ค่าคะแนนเฉลี่ย	57.87	57.87
ค่าเบี่ยงเบนมาตรฐาน	13.27	13.27
ค่าความคลาดเคลื่อนมาตรฐานค่าเฉลี่ย	0.871	0.871
คะแนนสูงสุด	90	90
คะแนนต่ำสุด	25	25
พิสัยของคะแนน	65	65
ค่าเบี่ยงเบนควอร์ไทล์	9.75	9.75
ค่ามัธยฐาน	57	57
ค่าความเบ้	-0.02	-0.02
ค่าความโค้ง	-0.45	-0.45

* $r_{xy} = 1.00$ ($p < 0.05$)

จากการเปรียบเทียบด้วยค่าสถิติพื้นฐานที่ได้จากการวิเคราะห์ข้อทดสอบราชข้อด้วยโปรแกรม Rasch กับโปรแกรม BICAL 7.0 จากข้อมูลชุดเดียวกันในตารางที่ 17 ปรากฏว่าทุกค่าในมิติต่างๆ เช่น ค่าคะแนนเฉลี่ย ค่าเบี่ยงเบนมาตรฐาน ค่าความเบ้ และค่าความโค้ง มีขนาดเท่ากันพอดี และเมื่อทำการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าสถิติต่างๆที่ได้จากโปรแกรมทั้งสองปรากฏว่า $r_{xy} = 1.00$

($p < 0.05$) หมายความว่าทุกค่ามีขนาดเท่ากันพอดี ซึ่งแสดงว่าโปรแกรม Rasch ทำการคำนวณได้ค่าสถิติพื้นฐานจากการวิเคราะห์ข้อทดสอบรายข้อแบบบราซได้ถูกต้องเช่นเดียวกับโปรแกรม BICAL 7.0

ตารางที่ 18: การเปรียบเทียบค่าความยาก (β) ความสอดคล้องกับโครงสร้างข้อทดสอบ และค่าความตรงรายข้อจากการวิเคราะห์ข้อทดสอบรายข้อ เมื่อคำนวณโดยโปรแกรม Rasch กับโปรแกรม BICAL 7.0 จากข้อมูลชุดเดียวกัน

Item N.	Names	จากโปรแกรม Rasch			จากโปรแกรม BICAL 7.0		
		β^*	Weighed MNSQ*	r_{pb}^*	β	Weighed MNSQ	r_{pb}
1	1	0.70	1.18	0.06	0.70	1.18	0.06
2	2	0.70	0.93	0.40	0.70	0.93	0.40
3	3	1.91	1.05	0.16	1.91	1.05	0.16
4	4	-1.33	1.01	0.22	-1.33	1.01	0.22
5	5	-1.62	0.82	0.49	-1.62	0.82	0.49
6	6	0.51	1.04	0.26	0.51	1.04	0.26
7	7	-1.08	0.99	0.24	-1.08	0.99	0.24
8	8	1.32	1.04	0.20	1.32	1.04	0.20
9	9	-0.31	1.00	0.29	-0.31	1.00	0.29
10	10	-1.79	0.93	0.33	-1.79	0.93	0.33

* $r_{xy} = 1.00$ ($p < 0.05$)

จากการเปรียบเทียบด้วยค่าเปล้าค่าความยาก (β) ความสอดคล้องกับโครงสร้างข้อทดสอบ และค่าความตรงรายข้อจากการวิเคราะห์ข้อทดสอบรายข้อด้วยโปรแกรม Rasch กับโปรแกรม BICAL 7.0 จากข้อมูลชุดเดียวกันในตารางที่ 18 ปรากฏว่าทุกค่าดังกล่าวในข้อที่ 1-10 ต่างมีขนาดเท่ากันพอดี และเมื่อทำการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าสถิติต่างๆที่ได้จากโปรแกรมทั้งสองปรากฏว่า $r_{xy} = 1.00$ ($p < 0.05$) หมายความว่าทุกค่ามีขนาดเท่ากันพอดี ซึ่งแสดงว่าโปรแกรม Rasch ทำการคำนวณได้ค่าความยาก (β) ความสอดคล้องกับโครงสร้างข้อทดสอบ และค่าความตรงรายข้อได้ถูกต้องเช่นเดียวกับโปรแกรม BICAL 7.0

ตารางที่ 19: การเปรียบเทียบค่าความสามารถของบุคคล (θ) และความคลาดเคลื่อนมาตรฐานของความสามารถจากการวิเคราะห์ข้อทดสอบรายข้อเมื่อคำนวณโดยโปรแกรม Rasch กับโปรแกรม BICAL 7.0 จากข้อมูลชุดเดียวกัน

Score	Count	จากโปรแกรม Rasch		จากโปรแกรม BICAL 7.0	
		θ	SE *	θ	SE
85	2	2.05	0.30	2.05	0.30
84	0	1.96	0.29	1.96	0.29
83	2	1.88	0.28	1.88	0.28
82	2	1.80	0.28	1.80	0.28
81	2	1.73	0.27	1.73	0.27
80	4	1.66	0.27	1.66	0.27
79	0	1.59	0.26	1.59	0.26
78	0	1.52	0.26	1.52	0.26
77	4	1.46	0.26	1.46	0.26
76	6	1.39	0.25	1.39	0.25
Separability Index		0.89		0.89	

* $r_{xy} = 1.00$ ($p < 0.05$)

จากการเปรียบเทียบด้วยค่าความสามารถของบุคคล (θ) และความคลาดเคลื่อนมาตรฐานของความสามารถจากการวิเคราะห์ข้อทดสอบรายข้อด้วยโปรแกรม Rasch กับโปรแกรม BICAL 7.0 จากข้อมูลชุดเดียวกันในตารางที่ 19 ปรากฏว่าทุกค่าดังกล่าวสำหรับผู้ที่ยอมรับได้คะแนนดิบระหว่าง 85-76 รวมทั้งค่าความเที่ยงของแบบทดสอบ (separability) ต่างมีขนาดเท่ากันพอดี และเมื่อทำการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าสถิติต่างๆที่ได้จากโปรแกรมทั้งสองปรากฏว่า $r_{xy} = 1.00$ ($p < 0.05$) หมายความว่าทุกค่ามีขนาดเท่ากันพอดี ซึ่งแสดงว่าโปรแกรม Rasch ทำการคำนวณค่าต่างๆดังกล่าวได้ถูกต้องเช่นเดียวกับโปรแกรม BICAL 7.0

ดังนั้น จากผลการเปรียบเทียบค่าสถิติต่างๆที่ได้จากการใช้โปรแกรม Rasch กับโปรแกรม BICAL 7.0 เพื่อการวิเคราะห์ข้อทดสอบรายข้อในตารางที่ 18-19 ข้างต้น ซึ่งได้ค่าเท่ากันพอดีทำให้สามารถสรุปได้ว่าโปรแกรม Rasch ทำงานได้ถูกต้องเช่นเดียวกับโปรแกรม BICAL 7.0

3. การเปรียบเทียบผลการทำงานของโปรแกรมธนาคารข้อทดสอบซึ่งเรียกชื่อว่าโปรแกรม Item Bank กับโปรแกรม ITB Pro

เพื่อทดสอบการทำงานของโปรแกรมธนาคารข้อทดสอบที่พัฒนาขึ้น (โปรแกรม Item Bank) ว่ามีความสามารถในการทำงานต่างๆที่เกี่ยวกับระบบธนาคารข้อทดสอบหรือแบบทดสอบได้ครบถ้วนหรือไม่ ผู้วิจัยได้นำข้อมูลจริงจากฐานข้อมูลที่ใช้กับโปรแกรมธนาคารข้อทดสอบเดิมที่เรียกชื่อว่า ITB Pro (Item Banking, Professional Version) มาทดสอบการทำงานของโปรแกรมวิเคราะห์ใหม่ที่พัฒนาขึ้น แล้วเปรียบเทียบที่เกิดขึ้นกับผลการทำงานของโปรแกรมเดิม ปรากฏผลดังนี้

ตารางที่ 20: การเปรียบเทียบความสามารถในการทำงานของโปรแกรม Item Bank ที่พัฒนาขึ้นใหม่กับโปรแกรม ITB Pro จากฐานข้อมูลชุดเดียวกัน

ความสามารถในการทำงาน	โปรแกรม Item Bank	โปรแกรม ITB Pro
1. การสร้างข้อทดสอบ	ทำได้ถูกต้อง	ทำได้
2. การแก้ไขข้อทดสอบ	ทำได้ถูกต้อง	ทำได้
3. การลบข้อทดสอบทิ้ง	ทำได้ถูกต้อง	ทำได้
4. การเพิ่มข้อทดสอบ	ทำได้ถูกต้อง	ทำได้
5. การเปลี่ยนรหัสข้อทดสอบ	ทำได้ถูกต้อง	ทำได้
6. การเก็บภาพประกอบ	ทำได้ถูกต้อง	ทำไม่ได้
7. การสุ่มเลือกข้อทดสอบ	ทำได้	ทำได้
8. การแสดงข้อทดสอบที่เลือกแล้ว	ทำได้ถูกต้อง	ทำได้
9. การจัดพิมพ์ข้อทดสอบ	ทำได้ถูกต้อง	ทำได้
10. การจัดรูปแบบของแบบทดสอบ	ทำได้	ทำได้
11. การสำรองฐานข้อมูล	ทำได้ถูกต้อง	ทำได้
12. การตรวจรหัสความปลอดภัยของฐานข้อมูล	ทำได้ถูกต้อง	ทำไม่ได้
13. การจัดการฐานข้อมูลตามอภิธานุกรม	ทำได้ซับซ้อนกว่า	ทำได้
3. 14. การใช้กับภาษาไทยและภาษาอังกฤษ	ทำได้ถูกต้อง	ทำได้
15. การจัดการต่างๆดังกล่าวกับแบบทดสอบ	ทำได้ถูกต้อง	ทำได้

จากการเปรียบเทียบความสามารถในการทำงานของโปรแกรม Item Bank กับโปรแกรม ITB Pro ด้วยฐานข้อมูลชุดเดียวกัน ปรากฏว่าโปรแกรมใหม่ที่พัฒนาขึ้นสามารถทำงานต่างๆที่เกี่ยวข้องกับการทำงานของธนาคารข้อทดสอบได้ดีเช่นเดียวกับโปรแกรมเดิม เช่น การเตรียมข้อทดสอบและแบบทดสอบชนิดต่างๆเข้าเก็บไว้ในฐานข้อมูล การปรับปรุงแก้ไข การคัดเลือกข้อทดสอบหรือแบบทดสอบ และการจัดพิมพ์สิ่งที่ได้เลือกแล้วในรูปแบบที่ต้องการได้ นอกจากนี้ยังพบว่า โปรแกรม Item Bank ยังสามารถทำงาน 3 อย่างได้ดีกว่าโปรแกรม ITB Pro คือ สามารถตรวจรหัสความปลอดภัยของฐานข้อมูล ได้ และจัดเก็บภาพประกอบข้อทดสอบหรือแบบทดสอบได้ ซึ่งความสามารถดังกล่าวนี้ไม่มีในโปรแกรม ITB Pro รวมทั้งสามารถจัดการเก็บและเลือกข้อมูลจากฐานข้อมูลที่แยกประเภทตามอนุกรมวิธาน (taxonomy) ได้ซับซ้อนมากกว่า

อนึ่ง ความถูกต้องในการทำงานของ โปรแกรม Item Bank เกี่ยวกับการสุ่มตัวอย่างข้อทดสอบหรือแบบทดสอบ และการจัดรูปแบบของแบบทดสอบไม่สามารถตรวจสอบโดยการเปรียบเทียบกับผลการทำงาน

ของโปรแกรม ITB Pro ได้โดยตรง เนื่องจากโปรแกรมทั้งสองทำงานในขั้นตอนดังกล่าวด้วยวิธีการสุ่มเลือกตัวอย่าง แต่สามารถอนุมานได้ว่า โปรแกรม Item Bank ทำงานได้ถูกต้องเพราะวิธีการสุ่มเป็นไปตามสูตรทางสถิติศาสตร์ที่สามารถพิสูจน์ได้

ดังนั้น จากผลการเปรียบเทียบความสามารถในการทำงานในด้านต่างๆ ที่ได้กล่าวมาแล้วข้างต้นจึงสรุปได้ว่า โปรแกรม Item Bank ทำงานได้ดีกว่า โปรแกรม ITB Pro ในการจัดการเกี่ยวกับธนาคารข้อทดสอบและแบบทดสอบ

4. การเปรียบเทียบผลการทำงานของโปรแกรมทดสอบแบบปรับเปลี่ยนที่พัฒนาขึ้น ซึ่งเรียกชื่อว่าโปรแกรม CCAT กับโปรแกรม CCAT Pro

เพื่อทดสอบการทำงานของโปรแกรมทดสอบแบบปรับเปลี่ยนที่พัฒนาขึ้น (โปรแกรม CCAT) ว่ามีความสามารถในการทำงานต่างๆ ที่เกี่ยวกับการทดสอบแบบปรับเปลี่ยน ได้ครบถ้วนหรือไม่ ผู้วิจัยได้นำข้อมูลจริงจากฐานข้อมูลที่ใช้กับโปรแกรมธนาคารข้อทดสอบเดิมที่เรียกชื่อว่า ITB Pro (Item Banking, Professional Version) มาทดสอบการทำงานของโปรแกรมวิเคราะห์ใหม่ที่พัฒนาขึ้น แล้วเปรียบเทียบที่เกิดขึ้นกับผลการทำงานของโปรแกรมเดิม (เนื่องจากโปรแกรมการทดสอบแบบปรับเปลี่ยนอาศัยข้อทดสอบจากฐานข้อมูลของโปรแกรมธนาคารข้อทดสอบ)ปรากฏผลดังนี้

ตารางที่ 21: การเปรียบเทียบความสามารถในการทำงานของโปรแกรม CCAT ที่พัฒนาขึ้นใหม่กับโปรแกรม CCAT Pro จากฐานข้อมูลชุดเดียวกัน

ความสามารถในการทำงาน	โปรแกรม CCAT	โปรแกรม CCAT Pro
1. การกำหนดปริเขตในการทดสอบ	ทำได้ถูกต้อง	ทำได้
2. การสุ่มเลือกข้อทดสอบ	ทำได้	ทำได้
3. การปรับเปลี่ยนข้อทดสอบตามเกณฑ์	ทำได้ถูกต้อง	ทำได้
4. การทำงานร่วมกับฐานข้อมูลธนาคารข้อทดสอบ	ทำได้ถูกต้อง	ทำได้
5. การคำนวณผลการทดสอบ	ทำได้ถูกต้อง	ทำได้
6. การรายงานวิธีการทดสอบ	ทำได้ถูกต้อง	ทำได้
7. การจัดเก็บผลการทดสอบ	ทำได้ถูกต้อง	ทำได้
8. การพิมพ์รายงานผลการทดสอบ	ทำได้ถูกต้อง	ทำได้

จากการเปรียบเทียบความสามารถในการทำงานของโปรแกรม CCAT กับโปรแกรม CCAT Pro ด้วยฐานข้อมูลชุดเดียวกัน ในตารางที่ 21 ปรากฏว่าโปรแกรมใหม่ที่พัฒนาขึ้นสามารถทำงานต่างๆที่เกี่ยวข้องกับการทดสอบแบบปรับเปลี่ยนได้ดีเช่นเดียวกับโปรแกรมเดิม เช่น การกำหนดปริเขตในการทดสอบ การสุ่มเลือกข้อทดสอบ การปรับเปลี่ยนข้อทดสอบตามเกณฑ์ และการรายงานวิธีการทดสอบ เป็นต้น

อนึ่ง ความถูกต้องในการทำงานของโปรแกรม CCAT เกี่ยวกับการสุ่มตัวอย่างข้อทดสอบมาให้ผู้สอบตอบไม่สามารถตรวจสอบโดยการเปรียบเทียบกับผลการทำงานของโปรแกรม ITB Pro ได้โดยตรง เนื่องจากโปรแกรมทั้งสองทำงานในขั้นนี้ด้วยวิธีการสุ่มเลือกตัวอย่าง แต่สามารถอนุมานได้ว่าโปรแกรม CCAT ทำงานได้ถูกต้องเพราะวิธีการสุ่มเป็นไปตามสูตรทางสถิติศาสตร์ที่สามารถพิสูจน์ได้

ดังนั้น จากผลการเปรียบเทียบความสามารถในการทำงานในด้านต่างๆที่ได้กล่าวมาแล้วข้างต้นจึงสรุปได้ว่าโปรแกรม CCAT ทำงานได้ดีเช่นเดียวกับโปรแกรม CCAT Pro ในการจัดการเกี่ยวกับการทดสอบแบบปรับเปลี่ยนตามความสามารถของผู้สอบ

5. การเปรียบเทียบผลการทำงานของโปรแกรมกำหนดระดับคะแนนที่พัฒนาขึ้น ซึ่งเรียกชื่อว่าโปรแกรม Grading and T-Score กับโปรแกรม Grading and NTScore

เพื่อทดสอบการทำงานของโปรแกรมกำหนดระดับคะแนนที่พัฒนาขึ้น (โปรแกรม Grading and T-Score) ว่ามีความสามารถในการทำงานต่างๆที่เกี่ยวกับการกำหนดระดับคะแนนด้วยวิธีต่างๆได้ถูกต้องหรือไม่ ผู้วิจัยได้นำข้อมูลจริงจากผลการทดสอบแบบทดสอบ CULI- TEPOIC ชุดที่ 1 มาตรวจคะแนนและวิเคราะห์รายชื่อด้วยโปรแกรม Classical ที่พัฒนาขึ้น และโปรแกรม CTIA 7.0 แล้วเลือกใช้โปรแกรม Grading and T-Score และโปรแกรม Grading and NTScore ทำการกำหนดระดับคะแนน (“ตัดเกรด”) โดยอาศัยเกณฑ์อย่างเดียวกันสำหรับแต่ละวิธี (เนื่องจากโปรแกรมกำหนดระดับคะแนนอาศัยข้อมูลผลการทดสอบจากฐานข้อมูลของโปรแกรมวิเคราะห์ข้อทดสอบแบบประเพณีนิยม) ยกเว้นวิธีใช้คะแนนมาตรฐานที่ (T-Score) ต้องป้อนข้อมูลชุดใหม่เพื่อให้โปรแกรมทำงานได้ ปรากฏผลการทดสอบดังนี้

ตารางที่ 22: การเปรียบเทียบจำนวนผู้ได้รับ“เกรด”จากการคำนวณแต่ละวิธีจากโปรแกรม Grading and T-Score กับ โปรแกรม Grading and NTScore จากข้อมูลชุดเดียวกัน

Grade	ผลจากโปรแกรม Grading and T-Score							ผลจากโปรแกรม Grading and NTScore						
	วิธีการกำหนดระดับคะแนน							วิธีการกำหนดระดับคะแนน						
	1*	2*	3*	4*	5*	6*	7*	1	2	3	4	5	6	7
A	12	12	7	9	4	24	12	12	12	7	9	4	24	12
B	24	24	18	30	27	12	21	24	24	18	30	27	12	21
C	45	45	26	42	52	44	44	45	45	26	42	52	44	44
D	23	23	32	30	29	15	22	23	23	32	30	29	15	22
F	12	12	33	5	4	15	17	12	12	33	5	4	15	17
total	116	116	116	116	116	116	116	116	116	116	116	116	116	116

* $r_{xy} = 1.00$ ($p < 0.05$)

จากการเปรียบเทียบจำนวนผู้ได้รับ“เกรด”จากการกำหนดระดับคะแนนวิธีเดียวกันจากโปรแกรม Grading and T-Score กับโปรแกรม Grading and NTScore ในตารางที่ 22 พบว่าเท่ากันทุกรายการ และทำให้ค่าสัมประสิทธิ์สหสัมพันธ์ของผลการกำหนดระดับคะแนนแต่ละวิธีมีค่าสมบูรณ์ คือ $r_{xy} = 1.00$ ($p < 0.05$) ดังนั้น จึงสรุปได้ว่าโปรแกรม Grading and T-Score สามารถทำงานได้ดีเช่นเดียวกับโปรแกรม Grading and NTScore ในการจัดการเกี่ยวกับการกำหนดระดับคะแนนสอบ

บทที่ 5

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

ความนำ

ปัจจุบันนี้เป็นยุคของเทคโนโลยีเพื่อการสื่อสารนานาชาติ (Information and Communication Technology Age) ที่มีความก้าวหน้าหลาย ๆ ด้าน เช่น วิทยาการคอมพิวเตอร์ การสื่อสารนานาชาติ การทำธุรกิจผ่านทางระบบ Internet และการเรียนรู้ผ่านทางระบบเครือข่ายคอมพิวเตอร์ (Internet) เป็นต้น นอกจากนี้ และคนทั่วโลกนิยมใช้ภาษาอังกฤษเป็นสื่อสากลเพื่อการสื่อสารนานาชาติหรือสืบค้นข้อมูลต่าง ๆ ที่เกี่ยวข้องทั้งทางระบบเครือข่ายคอมพิวเตอร์ หรือสื่ออื่น ๆ มากที่สุดในโลก กล่าวคือ ประมาณร้อยละ 30 ของผู้ใช้ระบบเครือข่ายคอมพิวเตอร์หรือประมาณ 312,924,679 ล้านคนใช้ภาษาอังกฤษ (Internet World Stat, 2006) นอกจากนี้ภาษาอังกฤษยังเป็นภาษาที่ใช้ในสิ่งพิมพ์มากที่สุดในโลก รวมทั้งเป็นภาษาซึ่งผู้ที่ไม่ใช่เจ้าของภาษาใช้เพื่อพูดเป็นจำนวนมากที่สุดในโลกด้วย (Vistawide, 2006) ด้วยเหตุนี้จึงเป็นที่ยอมรับกันทั่วไปด้วยว่าภาษาอังกฤษมีความจำเป็นต่อการศึกษาค้นคว้าของนิสิตนักศึกษาและคนทั่วไป รวมทั้งมีความสำคัญต่อการประกอบธุรกิจของนักธุรกิจทั่วไปและการประกอบอาชีพของผู้ที่อยู่ในวงการอื่น ๆ อีก

มหาวิทยาลัยมีหน้าที่และความรับผิดชอบโดยตรงต่อสังคมในการช่วยสร้างบัณฑิตให้มีความรู้และความสามารถทันต่อการเปลี่ยนแปลงของสังคม และจากการวิจัยเชิงสำรวจจำนวนมาก มักพบว่านักศึกษา บัณฑิตและหน่วยงานต่าง ๆ ของราชการ องค์กรของรัฐวิสาหกิจ และบริษัทห้างร้านเอกชน มีความต้องการและมีการใช้ภาษาอังกฤษในการบริหารกิจการในระดับปานกลางถึงในระดับสูง โดยเฉพาะอย่างยิ่งผู้ที่เป็นนักวิชาการและบุคคลในระดับบริหารกิจการต้องใช้ภาษาอังกฤษเพื่อการอ่าน เขียน ฟัง และพูดในระดับมากตามลำดับ ปริมาณความต้องการเช่นนี้มีมานานแล้ว และมีแนวโน้มว่าจะมีมากขึ้นเรื่อย ๆ (อังฉรา วงศ์โสธร และคณะ, 2523:20; บังอร สว่างวโรรส, 2527:38 สุพัฒน์ สุขมลสันต์ และคณะ, 2529:63-68)

ดังนั้น ในปี พ.ศ. 2533 จุฬาลงกรณ์มหาวิทยาลัย ได้เล็งเห็นความจำเป็นในการสื่อสารนานาชาติระหว่างประเทศโดยใช้ภาษาอังกฤษเป็นสื่อกลางเป็นอย่างมาก โดยเชื่อว่า หากบัณฑิตของจุฬาลงกรณ์มหาวิทยาลัยมีความสามารถทั้งทางเนื้อหาวิชาการ และการใช้ภาษาอังกฤษในระดับที่สามารถติดต่อทำความเข้าใจกับชาวต่างประเทศได้ ก็จะทำให้บัณฑิตมีคุณค่าเป็นที่ต้องการของตลาดแรงงานและสังคมมากขึ้น ซึ่งจะทำให้เป็นแรงผลักดันให้สังคมไปสู่สังคมอุตสาหกรรมได้ทันเวลาและปลอดภัย ดังนั้น ฝ่ายวิชาการของมหาวิทยาลัยจึงเห็นความจำเป็นที่จะต้องปรับความสามารถในการใช้ภาษาอังกฤษของบัณฑิตให้ได้มาตรฐานที่ทางมหาวิทยาลัยกำหนด และวิธีหนึ่งที่จะทำได้โดยการทดสอบสมิทธิภาพทั่วไปทางภาษา อังกฤษของนิสิตชั้นปีที่ 3-4 หรือ 2 ปีสุดท้ายของหลักสูตรต่าง ๆ เพื่อเป็นการกระตุ้นให้เกิดการ

คืนตัวทั้งอาจารย์และนิสิตให้สนใจและให้ความสำคัญต่อการเรียนการสอนภาษาอังกฤษด้วยวิธีการต่าง ๆ (อัจฉรา และคณะ, 2535; สุพัฒน์ และคณะ, 2533) ดังนั้นในปี พ.ศ. 2533-2536 ฝ่ายวิชาการของมหาวิทยาลัย จึงได้ร่วมกับฝ่ายวิจัยและพัฒนาของสถาบันภาษา ทำการวิจัยและพัฒนาแบบทดสอบวัดสมรรถภาพทั่วไปทางภาษาอังกฤษของจุฬาลงกรณ์มหาวิทยาลัย (Chulalongkorn University Test of English Proficiency หรือ CU-TEP) ขึ้นตามแนวแบบทดสอบ TOEFL (Test of English as a Foreign Language) (สุพัฒน์ และคณะ, 2536) และได้มีการสร้างและพัฒนาแบบทดสอบดังกล่าวติดต่อกันเรื่อยมา เพื่อให้บริการการทดสอบแก่นิสิต นักศึกษา และบุคคลทั่วไปที่สนใจทั้งภาครัฐ รัฐวิสาหกิจและเอกชน จนเป็นที่ยอมรับกันทั่วไปในวงการศึกษาระดับอุดมศึกษาและวงการธุรกิจบางสาขา และทำให้มหาวิทยาลัยจัดตั้งศูนย์การทดสอบแห่งจุฬาลงกรณ์มหาวิทยาลัย (Testing Center of Chulalongkorn University) ขึ้น

แต่อย่างไรก็ตาม เนื่องจากแบบทดสอบ CU-TEP วัดสมรรถภาพทั่วไปทางภาษาอังกฤษสำหรับผู้ที่ต้องการวัดความสามารถทั่วไปทางภาษาอังกฤษของตนเอง จึงยังไม่เหมาะสำหรับการทดสอบความสามารถเฉพาะทางเพื่อวิชาชีพและการสื่อสารนานาชาติ ซึ่งเป็นแบบทดสอบที่สังคมมีความต้องการมากขึ้น ดังจะเห็นได้จากมีแบบทดสอบที่นิยมใช้ทั่วไปขณะนี้ คือแบบทดสอบ TOEIC (Test of English for International Communication) ของ ETS (Educational Testing Service) แห่งอเมริกา แม้ว่าค่าบริการทดสอบของแบบทดสอบนี้ยังสูงมากสำหรับคนทั่วไป

ดังนั้น สถาบันภาษาจึงเห็นความสำคัญที่จะจัดให้มีโครงการวิจัยเรื่อง การพัฒนาแบบทดสอบวัดสมรรถภาพทางภาษาอังกฤษเพื่อวิชาชีพและการสื่อสารนานาชาติของสถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย หรือ CULI-TEPOIC [Chulalongkorn University Language Institute Test of English Proficiency for Occupational and International Communication Purposes] ให้ได้มาตรฐานขึ้น และทำการพัฒนาโปรแกรมคอมพิวเตอร์ที่จำเป็นต้องใช้เพื่อการพัฒนาข้อทดสอบและการทดสอบให้ทันสมัยยิ่งขึ้น ทั้งนี้เพื่อให้ระบบการทดสอบของสถาบันภาษามีความสมบูรณ์ยิ่งขึ้นด้วย

วัตถุประสงค์ของโครงการวิจัย

1. เพื่อสร้างและพัฒนาแบบทดสอบวัดสมรรถภาพทางภาษาอังกฤษเพื่อวิชาชีพและการสื่อสารนานาชาติของสถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย ให้มีคุณภาพและได้มาตรฐาน จำนวน 3 ชุด
2. เพื่อพัฒนาโปรแกรมคอมพิวเตอร์ที่ใช้ในการพัฒนาแบบทดสอบและการทดสอบให้ทันสมัย 5 โปรแกรม

ขอบข่ายของการวิจัย

1. แบบทดสอบที่สร้างและพัฒนาเป็นแบบทดสอบวัดสมรรถภาพทางภาษาอังกฤษเพื่อวิชาชีพและการสื่อสารนานาชาตินานาชาติจำนวน 3 ชุดๆละ 100 ข้อในลักษณะที่คล้ายคลึงกับแบบทดสอบ TOEIC กล่าวคือแบบทดสอบแต่ละชุดประกอบด้วยดังนี้

ก. การฟัง (Listening) จำนวน 50 ข้อ ประกอบด้วย

- Part 1: ฟังข้อความประกอบรูปภาพ (Photographs) จำนวน 10 ข้อ
- Part 2: ฟังคำถามและตอบ (Question – Response) จำนวน 10 ข้อ
- Part 3: ฟังข่าวและประกาศ (News and Announcements) จำนวน 10 ข้อ
- Part 4: ฟังบทสนทนาสั้นๆ (Short Conversations) จำนวน 10 ข้อ
- Part 5: ฟังการพูดสั้นๆ (Short Talks) จำนวน 10 ข้อ

ข. การเขียนและอ่าน (Writing and Reading) จำนวน 50 ข้อ ประกอบด้วย

- Part 6: อ่านและเติมคำในประโยค (Incomplete Sentences) จำนวน 15 ข้อ
- Part 7: อ่านหาข้อผิดพลาด (Error Recognition) จำนวน 15 ข้อ
- Part 8: อ่านเข้าใจความ (Reading Comprehension) จำนวน 20 ข้อ

ค. ระยะเวลาในการทดสอบทั้งหมดรวม 1.40 ชั่วโมง คือ

การฟัง (ก) ใช้เวลา 40 นาที

การอ่าน (ข) ใช้เวลา 60 นาที

2. แบบทดสอบที่สร้างและพัฒนาเป็นแบบทดสอบสำหรับนิสิต/นักศึกษาในระดับมหาวิทยาลัยและบุคคลทั่วไปที่สนใจที่เน้นเรื่องการสื่อสารนานาชาติเพื่อวิชาชีพและการสื่อสารนานาชาติ
3. โปรแกรมคอมพิวเตอร์ที่สร้างขึ้นเพื่อใช้ในการพัฒนาแบบทดสอบและการทดสอบมีจำนวน 5 โปรแกรม คือ โปรแกรมวิเคราะห์ข้อทดสอบแบบประเพณีนิยม (Classical Model) โปรแกรมวิเคราะห์ข้อทดสอบแบบราช (Rasch Model) โปรแกรมธนาคารข้อทดสอบ (Item Banking) โปรแกรมการทดสอบแบบปรับเปลี่ยน (Adaptive Testing) และโปรแกรมกำหนดระดับคะแนน (Grade Assignment)

4. โปรแกรมคอมพิวเตอร์ที่สร้างขึ้นทั้งหมดเป็นการพัฒนามาจากโปรแกรมเดิมที่ใช้ในระบบ DOS (Disk Operation System) เป็นระบบ Windows (Window Operation System)

ข้อตกลงเบื้องต้น

ในการวิจัยครั้งนี้ ผู้วิจัยมีข้อตกลงเบื้องต้น (หรือข้อสันนิษฐานเบื้องต้น) ที่เชื่อว่าเป็นจริงดังต่อไปนี้คือ

1. ผู้สอบแบบทดสอบที่ใช้ในการวิจัยครั้งนี้ได้พยายามมากที่สุดในการตอบแบบทดสอบเพื่อให้ได้คะแนนมากที่สุดเท่าที่จะสามารถทำได้ เพราะการทดสอบแต่ละครั้งเป็นการทดสอบจริงเพื่อทดสอบความรู้และความสามารถของตนเองว่ามีมากน้อยเพียงใด หรือนำผลการทดสอบไปประกอบการพิจารณาสมัครเข้าทำงานของตนเอง ซึ่งเป็นความต้องการของผู้สอบแต่ละคน
2. โปรแกรมคอมพิวเตอร์ทั้ง 5 โปรแกรมที่ใช้ในระบบ DOS เป็นโปรแกรมที่ทำงานได้ผลการคำนวณถูกต้องและเชื่อถือได้ เนื่องจากเป็นโปรแกรมที่ได้พัฒนาและตรวจสอบผลการทำงานมานานแล้วว่าถูกต้องและเชื่อถือได้ (แต่ที่ต้องการปรับปรุงและพัฒนาให้สามารถทำงานได้ในระบบ Windows เนื่องจากปัจจุบันนี้เครื่องคอมพิวเตอร์โดยทั่วไปไม่ได้ใช้ระบบ DOS แล้ว)

ความจำกัดของการวิจัย

การวิจัยนี้มีข้อจำกัดบางอย่างดังนี้

1. แบบทดสอบที่ใช้ในการวิจัยครั้งนี้มี 3 ชุด คือ CULI-TEPOIC ชุด 1 ชุด 2 และชุด 3 แบบทดสอบทั้ง 3 ชุดนี้ สร้างจากตารางกำหนดลักษณะเฉพาะของข้อทดสอบ (Table of Test Specification) เดียวกัน แต่ทำการทดสอบกับกลุ่มผู้สอบที่แตกต่างกัน แม้ว่าจะอยู่ในระดับอุดมศึกษาและบุคคลทั่วไปเช่นเดียวกันก็ตาม ดังนั้น ค่าสถิติที่ได้จากการวิเคราะห์ข้อทดสอบแต่ละชุดอาจแตกต่างกันบ้าง
2. การคำนวณหาความตรงร่วมสมัย (concurrent validity) ของแบบทดสอบที่สร้างขึ้นทั้ง 3 ชุดอาศัยคะแนนผลการสอบจากแบบทดสอบ TOEIC กับคะแนนผลการสอบจากแบบทดสอบที่สร้างขึ้น ผู้วิจัยได้สุ่มเลือกผู้ที่สอบได้คะแนนในระดับต่างๆอย่างมีระบบจำนวนชุดละ 25 คนเพื่อให้ไปสอบแบบทดสอบ TOEIC แต่ปรากฏว่ามีผู้ไปสอบเพียงชุดละ 19-24 คนเท่านั้น ซึ่งเป็นจำนวนที่ค่อนข้างน้อย ดังนั้น สมการพยากรณ์คะแนนผลการทดสอบแบบทดสอบ TOEIC จึงยังมีความคลาดเคลื่อนค่อนข้างมาก

คำนิยามศัพท์ที่ใช้ในการวิจัยครั้งนี้

1. **นิสิต/นักศึกษาไทย** หมายถึงผู้สอบซึ่งเป็นคนไทยที่มีความรู้ความสามารถในระดับอุดมศึกษาและสมัครสอบแบบทดสอบวัดสมรรถภาพทางภาษาอังกฤษเพื่อวิชาชีพและการสื่อสารนานาชาติของสถาบันภาษาจุฬาลงกรณ์มหาวิทยาลัย หรือ CULI-TEPOIC [Chulalongkorn University Language Institute Test of English Proficiency for Occupational and International Communication Purposes] เพื่อการศึกษาต่อในประเทศหรือต่างประเทศ บุคคลเหล่านี้มีระดับความสามารถทางภาษาอังกฤษแตกต่างกัน และมีภูมิหลังทางการศึกษา สังคม และเศรษฐกิจต่างกัน
2. **แบบทดสอบสมรรถภาพ (Proficiency Test)** หมายถึงแบบทดสอบที่ประกอบด้วยข้อทดสอบที่มุ่งทดสอบความรู้ความสามารถทั่วไปที่จำเป็นต่อการกระทำกรอย่างใดอย่างหนึ่งในอนาคต เช่น เพื่อการศึกษาต่อ เพื่อการสื่อสารนานาชาติ และเพื่อใช้ในวิชาชีพ เป็นต้น และเนื้อหาเหล่านี้ไม่ได้กำหนดไว้ในหลักสูตรของรายวิชาใดวิธีหนึ่งโดยเฉพาะ
3. **สมรรถภาพทั่วไปทางภาษาอังกฤษเพื่อวิชาชีพและการสื่อสารนานาชาติ** หมายถึงความรู้ความสามารถของผู้สอบที่ได้จากคะแนนสอบของแบบทดสอบวัดสมรรถภาพทางภาษาอังกฤษเพื่อวิชาชีพและการสื่อสารนานาชาติของสถาบันภาษาจุฬาลงกรณ์มหาวิทยาลัย หรือ CULI-TEPOIC [Chulalongkorn University Language Institute Test of English Proficiency for Occupational and International Communication Purposes] คะแนนเหล่านี้บ่งบอกถึงความรู้ความสามารถทั่วไปทางภาษาอังกฤษของผู้สอบที่ไม่ได้ระบุไว้ในรายวิชาใดวิชาหนึ่งโดยเฉพาะ แต่มีความจำเป็นต่อการสื่อสารนานาชาติในวิชาชีพและการสื่อสารนานาชาติทั่วไป
4. **โปรแกรมคอมพิวเตอร์ที่ใช้ในการพัฒนาแบบทดสอบและการทดสอบ** หมายถึงโปรแกรมวิเคราะห์ข้อทดสอบแบบประเพณีนิยม (Classical Model) โปรแกรมวิเคราะห์ข้อทดสอบแบบราช (Rasch Model) โปรแกรมธนาคารข้อทดสอบ (Item Banking) โปรแกรมการทดสอบแบบปรับเปลี่ยน (Adaptive Testing) และโปรแกรมกำหนดระดับคะแนน (Grade Assignment)

ประโยชน์ที่คาดว่าจะได้รับ และหน่วยงานที่นำผลการวิจัยไปใช้ประโยชน์

1. นิสิต/นักศึกษาและบุคคลทั่วไปเมื่อทำการทดสอบแล้วจะได้ทราบว่าจะระดับสมรรถภาพทั่วไปเพื่อวิชาชีพและการสื่อสารนานาชาติทางภาษาอังกฤษของตนมีมากน้อยเพียงใด ซึ่งจะเป็นตัวกระตุ้นให้นิสิต/นักศึกษาและบุคคลทั่วไปมีความสนใจ หรือ

- แรงจูงใจในการเรียนภาษาอังกฤษมากยิ่งขึ้น และเป็นผลดีต่อสภาพการเรียนการสอน ภาษาอังกฤษในประเทศได้ทางหนึ่ง
2. สถาบันต่าง ๆ หน่วยงานของรัฐ และบริษัทห้างร้านต่าง ๆ ของเอกชนอาจใช้คะแนนผลการทดสอบประกอบการพิจารณาเพื่อคัดเลือกนิสิต/นักศึกษาและบุคคลทั่วไปรับเข้าทำงานในหน่วยงานของตน โดยไม่ต้องเสียค่าใช้จ่ายในการพัฒนาแบบทดสอบของตนเองให้สิ้นเปลืองงบประมาณ
 3. คณะต่างๆของมหาวิทยาลัยอาจพิจารณาเกณฑ์ความสามารถจากระดับคะแนนสอบเป็นเกณฑ์ความ สามารถขั้นต่ำของนิสิต/นักศึกษาของตนก่อนที่จะจบการศึกษา เพื่อเป็นการควบคุมคุณภาพทางภาษาของบัณฑิตของตนได้
 4. ประชาชาติจะสามารถประหยัดเงินตราของประเทศได้ปีละจำนวนมาก หากว่าเกณฑ์คะแนนสอบได้รับการยอมรับทั่วไป เพราะนิสิต/นักศึกษาและบุคคลทั่วไปไม่จำเป็นต้องไปทดสอบแบบทดสอบอื่นของต่างประเทศที่ต้องเสียค่าสมัครสอบแพงมาก และในขณะเดียวกันก็เป็นการส่งเสริมการใช้ภูมิปัญญาไทยในการประหยัดเงินตราของประเทศได้ด้วย
 5. สถาบันภาษาจะได้แบบทดสอบที่มีคุณภาพและได้มาตรฐานจำนวน 3 ชุด หรือ 300 ข้อ เพื่อเก็บไว้ใช้ในธนาคารข้อทดสอบสำหรับใช้ทดสอบวัดสมิทธิภาพทางภาษาอังกฤษเพื่อวิชาชีพและการสื่อสารนานาชาติในโอกาสต่าง ๆ ที่ต้องการทันที
 6. สถาบันภาษาสามารถให้บริการทดสอบวัดสมิทธิภาพทางภาษาอังกฤษให้แก่หน่วยงานและองค์กรต่าง ๆ ที่ขอรับบริการได้ทันที ทำให้งานบริการทางวิชาการด้านทดสอบมีประสิทธิภาพและสนองความต้องการของผู้รับบริการได้มากยิ่งขึ้นนอกเหนือจากการทดสอบวัดสมิทธิภาพทั่วไปทางภาษาอังกฤษอื่นที่มีอยู่แล้ว
 7. สถาบันภาษาจะได้โปรแกรมคอมพิวเตอร์จำนวน 5 โปรแกรม ซึ่งทันสมัยสำหรับใช้ในระบบการทดสอบ โดยไม่ต้องเสียเงินไปซื้อจากต่างประเทศด้วยราคาที่แพงมาก และอาจเป็นประโยชน์กับหน่วยงาน หรือองค์กรอื่นที่ดำเนินการด้านการทดสอบ หรือการให้บริการทดสอบได้ด้วย

วิธีการวิจัย

ส่วนที่ 1 การสร้างและพัฒนาแบบทดสอบ CULI-TEPOIC

ประชากร (population)

ประชากรของการวิจัยครั้งนี้ ได้แก่ นิสิต นักศึกษา และบุคคลภายนอกที่สนใจสอบแบบทดสอบวัดสมิทธิภาพทางภาษาอังกฤษเพื่อวิชาชีพและการสื่อสารนานาชาติของสถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย หรือ CULI-TEPOIC [Chulalongkorn University Language Institute Test of English

Proficiency for Occupational and International Communication Purposes] ในจุฬาลงกรณ์มหาวิทยาลัย และภายนอก ที่สถาบันภาษา จุฬาลงกรณ์ มหาวิทยาลัยได้ประชาสัมพันธ์ให้ใช้บริการทดสอบ โดยไม่ต้อง เสียค่าธรรมเนียมสอบ และผู้เข้าสอบบางคนมีโอกาสดำเนินการสอบแบบทดสอบ TOEIC (Test of English for International Communication) ของ ETS (Educational Testing Service) อีกด้วย ดังนั้นจึงไม่สามารถทราบจำนวนประชากรที่แน่นอนได้

พลวิจัย (subjects)

พลวิจัยของการวิจัยครั้งนี้ ได้แก่ นิสิต นักศึกษา และบุคคลภายนอกที่สนใจสอบแบบทดสอบ CULI-TEPOIC ชุดที่ 1, 2 และ 3 บุคคลเหล่านี้ มีความรู้ความสามารถทางภาษาอังกฤษหลากหลาย และสมัครมาสอบเองด้วยวัตถุประสงค์ที่แตกต่างกัน เช่น เพื่อเตรียมตัวสอบชิงทุนไปศึกษาต่อ ณ ต่างประเทศ เพื่อเตรียมตัวสอบคัดเลือกไปดูงาน ณ ต่างประเทศ บางคนสอบเพื่อเตรียมตัวสอบเลื่อนขั้นเงินเดือน หรือ ตำแหน่ง บางคนมาสอบเพื่อให้ผ่านเกณฑ์ระดับความสามารถทางภาษาตามที่หน่วยงานกำหนดไว้ และ บางคนมาสอบเพื่อทดสอบความรู้ทั่วไปทางภาษาอังกฤษของตนเอง เป็นต้น การทดสอบด้วย วัตถุประสงค์ต่าง ๆ เหล่านี้ มีผู้เข้าสอบ จำนวนดังต่อไปนี้

ตารางที่ 23: จำนวนพลวิจัย

ที่	ชื่อแบบทดสอบ	จำนวนผู้สอบ
1	CULI-TEPOIC ชุด 1	116
2	CULI-TEPOIC ชุด 2	124
3	CULI-TEPOIC ชุด 3	120
	รวมทั้งสิ้น	360

เครื่องมือการวิจัย

ก. ลักษณะทั่วไป

เครื่องมือการวิจัยครั้งนี้ ได้แก่ แบบทดสอบ CULI-TEPOIC จำนวน 3 ชุด คือ ชุด 1, 2 และ 3 แต่ละชุดเป็นข้อทดสอบปรนัยแบบเลือกคำตอบ 4 ตัวเลือก และมี 2 ส่วน รวมเป็น 100 ข้อ คือ

1. ข้อทดสอบการฟังเข้าใจความ 50 ข้อ
2. ข้อทดสอบการเขียนและอ่านเข้าใจความ 50 ข้อ

ข. การสร้างและการพัฒนา

แบบทดสอบทั้ง 3 ชุด เป็นแบบทดสอบที่จัดสร้างและพัฒนาขึ้นโดยคณะกรรมการสร้างและพัฒนาแบบทดสอบ CULI-TEPOIC ของสถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย ในปี 2546 โดยการอิง

แนวคิดและรูปแบบของแบบทดสอบ TOEIC แต่มีวัตถุประสงค์ในการทดสอบแตกต่างกัน กล่าวคือ แบบทดสอบ CULI-TEPOIC สร้างขึ้นเพื่อใช้ทดสอบสมรรถภาพทั่วไปทางภาษาอังกฤษเพื่อวิชาชีพและการสื่อสารนานาชาติ แต่แบบทดสอบ TOEIC สร้างขึ้นเพื่อใช้ทดสอบสมรรถภาพทั่วไปทางภาษาอังกฤษเพื่อการสื่อสารนานาชาติทั่วไป ขั้นตอนการสร้างและพัฒนาแบบทดสอบดังกล่าวมีรายละเอียดดังต่อไปนี้

1. เสนอแต่งตั้งคณะกรรมการสร้างและพัฒนาแบบทดสอบ CULI-TEPOIC ของสถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย ในปี 2546 จำนวน 7 คน เพื่อสร้างและพัฒนาแบบทดสอบ CULI-TEPOIC จำนวน 3 ชุด
2. คณะกรรมการฯ ศึกษาวัตถุประสงค์และรูปแบบของแบบทดสอบ TOEIC เพื่อใช้เป็นแนวทางในการสร้างแบบทดสอบที่ต้องการ
3. คณะกรรมการฯ ศึกษาหนังสือและตำราในการสร้างแบบทดสอบภาษาอังกฤษต่างๆ เช่น Alderson (2000), Bachman (1990), Bachman and Palmer (1996), Bloom and Others (1956), Brown (2001), Canale and Swain (1980), Canale (1983), Dunkel (1999), Osterlind (1989) และ Valette (1969) เป็นต้น
4. ผู้วิจัยทำการศึกษาคู่มือและขั้นตอนในการสร้างและพัฒนาแบบทดสอบให้เป็นแบบทดสอบมาตรฐานจากหนังสือ บทความ และตำราต่างๆ เช่น Devine and Yaghlian (2000), Anderson (1972), Zimmerman and Others (1990), Miller and Sundre (2000) และ Burton and Others (1991) เป็นต้น ซึ่งสรุปได้เป็น 7 ขั้นตอนดังนี้ คือ
 1. สร้างตารางลักษณะข้อทดสอบที่ต้องการเพื่อกำหนดสัดส่วนของข้อทดสอบสำหรับแต่ละวัตถุประสงค์
 2. สร้างข้อทดสอบตามลักษณะข้อทดสอบที่กำหนดไว้จำนวนหนึ่ง
 3. ให้ผู้เชี่ยวชาญทางภาษาตรวจสอบความถูกต้องของภาษาและความตรงเชิงเนื้อหาของข้อทดสอบ
 4. ทดลองสอบข้อทดสอบกับกลุ่มตัวอย่างที่เป็นตัวแทนของประชากรเป้าหมาย เมื่อวันที่ 20 กรกฎาคม 2546
 5. ทำการวิเคราะห์ข้อทดสอบรายข้อเพื่อตัดสินใจว่าควรปรับปรุงแก้ไขหรือตัดข้อทดสอบบางข้อทิ้ง
 6. ทดลองสอบข้อทดสอบที่ปรับปรุงแล้วอีกครั้งหนึ่ง
 7. กำหนดค่าความเที่ยงและความตรงของแบบทดสอบ
5. ดำเนินการสร้างและพัฒนาแบบทดสอบตามขั้นตอนต่างๆดังกล่าวแล้ว ยกเว้นขั้นตอนที่ 6 เนื่องจากการวิจัยครั้งนี้พบว่า มีผลวิจัยในการทดสอบแบบทดสอบแต่ละชุดจำนวนมากพอที่จะทำหาค่าสถิติในขั้นที่ 5 คงที่เพียงพอ และค่าสถิติต่างๆที่ได้จากการวิเคราะห์ข้อทดสอบ

รายชื่อ รวมทั้งแบบทดสอบทั้งฉบับสูงมากพอที่จะสามารถนำแบบทดสอบไปใช้ต่อไปได้ นอกจากนี้ยังพบว่า มีข้อทดสอบที่ควรต้องปรับปรุงแก้ไขจำนวนไม่มากนัก สามารถที่จะปรับปรุงให้ดีขึ้นแล้วนำไปใช้สำหรับการทดสอบจริงได้ในอนาคต

ดังนั้น ในขั้นที่ 6 จึงเป็นการสุ่มผู้เข้าสอบแบบทดสอบแต่ละชุดประมาณ 40 คน เพื่อไปสอบแบบทดสอบ TOEIC เมื่อวันที่ 23 สิงหาคม 2546 เพื่อนำคะแนนผลการทดสอบมา คำนวณหาค่าความตรงร่วมสมัย (Concurrent validity) และสร้างตารางพยากรณ์ผลการทดสอบแบบทดสอบ TOEIC จากผลการทดสอบแบบทดสอบ CULI-TEPOIC แทนที่จะทำการทดลองสอบแบบทดสอบอีกครั้งหนึ่ง

การวิเคราะห์ข้อมูล

1. ทำการวิเคราะห์ข้อทดสอบของแบบทดสอบแต่ละชุด ด้วยวิธีประเพณีนิยม (classical model) เพื่อหาคุณภาพของแบบทดสอบรายชื่อและทั้งฉบับ โดยใช้โปรแกรมคอมพิวเตอร์ชื่อ CTIA/Grading (สุพัฒน์ สุกมลสันต์, 2542: 30-65)
2. ทำการทดสอบหาความเป็นเอกมิติ (unidimensionality) ของแบบทดสอบทั้ง 3 ชุด โดยการวิเคราะห์องค์ประกอบของแบบทดสอบโดยใช้โปรแกรม SPSS/PC แล้วนำค่า factor loading มาหาค่าสัมประสิทธิ์สหสัมพันธ์กับค่า Biserial Correlation ซึ่งได้จากขั้นตอนที่ 1 และได้ค่า $r_{xy} = 0.87-0.93$ แสดงว่าแบบทดสอบแต่ละชุดมีความเป็นเอกมิติ (Warm, 1978: 101)
3. วิเคราะห์ข้อทดสอบของแบบทดสอบแต่ละชุด ด้วยวิธีโลจิสต์แบบราช (Logist/Rasch model) เพื่อวิเคราะห์หาค่าความยาก (β) ที่ไม่เปลี่ยนแปลงของข้อทดสอบ ทั้งนี้เพื่อใช้ประกอบการตัดสินใจในการเรียงลำดับข้อทดสอบตามความยากง่าย โดยใช้โปรแกรมคอมพิวเตอร์ชื่อ Rasch (สุพัฒน์ สุกมลสันต์, 2542: 67-84)
4. นำผลการวิเคราะห์ในขั้นที่ 1 มาศึกษาข้อดี และข้อบกพร่องที่ควรปรับปรุงแก้ไขข้อทดสอบรายชื่อให้มีคุณภาพดียิ่งขึ้น โดยอาศัยเกณฑ์การคัดเลือกข้อทดสอบปรนัยแบบเลือกตอบชนิด 4 ตัวเลือกที่เป็นแบบทดสอบสมิททิภาพทั่วไปดังนี้ (Cheung and Bucat, 2002; DIIA, 2003; Runte, 2001; Jerard, 1995; สุพัฒน์ สุกมลสันต์, 2542; Ebel, 1979)

ก. ตัวเลือกที่ถูกต้อง

1. มีค่าดัชนีความยาก (difficulty index: p) = 0.15 – 0.85
2. มีค่าดัชนีอำนาจจำแนก (discrimination index: d) = > 0.20
3. มีค่าดัชนีความตรงรายชื่อ (item validity index: r_{pb}) = > 0.20 หรือ มีนัยสำคัญ

ข. ตัวเลือกที่ผิด

1. มีค่าดัชนีความยาก (difficulty index: p) = 0.15 – 0.85
2. มีค่าดัชนีอำนาจจำแนก (discrimination index: d) = > 0.20 และคิดลบ

- มีค่าดัชนีความตรงรายข้อ (item validity index: r_{pb}) = > 0.20 และติลลป หรือไม่มีนัยสำคัญ

ก. ตัวเลือกที่ควรปรับปรุง

- มีค่าดัชนีความยาก (difficulty index: p) ของข้อที่ถูก < 0.15 หรือ > 0.85
- ตัวเลือกที่ผิดมีอัตราส่วนของดัชนีความยากแตกต่างจากตัวเลือกผิดอื่นมาก
- มีค่าดัชนีอำนาจจำแนก (discrimination index: d) < 0.20 หรือมีค่าเป็นบวก
- ตัวเลือกที่ผิดมีอัตราส่วนของดัชนีอำนาจจำแนกแตกต่างจากตัวเลือกผิดอื่นมาก
- มีค่าดัชนีความตรงรายข้อ (item validity index: r_{pb}) < 0.25 หรือมีค่าเป็นบวก
- ตัวเลือกที่ผิดมีอัตราส่วนของดัชนีความตรงรายข้อแตกต่างจากตัวเลือกผิดอื่นมาก
- ข้อที่ตัวเลือกที่ถูกมีค่าเฉลี่ยของคะแนนเกณฑ์ (Mean criterion score) น้อยกว่าของตัวเลือกที่ผิด

ง. ดัชนีความเที่ยง (reliability index)

- แบบ KR20 หรือ Cronbach alpha = 0.70 – 0.80 หมายความว่าแบบทดสอบมีความเที่ยงในระดับดี
 - แบบ KR20 หรือ Cronbach alpha = 0.81 – 0.90 หมายความว่าแบบทดสอบมีความเที่ยงในระดับดีมาก
 - แบบ KR20 หรือ Cronbach alpha > 0.91 หมายความว่าแบบทดสอบมีความเที่ยงในระดับดียอดเยี่ยม
 - แบบ KR20 หรือ Cronbach alpha < 0.60 หมายความว่าแบบทดสอบมีความเที่ยงต่ำและควรปรับปรุงแก้ไข เว้นแต่มีข้อทดสอบจำนวนน้อย เช่น 10 เป็นต้น
5. คัดเลือกข้อทดสอบที่ดีและข้อที่ควรปรับปรุงแก้ไขจากแบบทดสอบทั้ง 3 ฉบับ เพื่อนำไปแก้ไขสำหรับใช้ในการทดสอบจริงต่อไป

สถิติที่ใช้

1. ความยากง่ายของข้อทดสอบแบบประเพณีนิยม (สุพัตน์ สุขมลสันต์, 2538:18-19)

$$p = \frac{N_p}{N_i}$$

เมื่อ N_p = จำนวนผู้สอบทั้งหมดที่เลือกตอบตัวเลือก i

N_i = จำนวนผู้เข้าสอบทั้งหมด

2. อำนาจจำแนกของข้อทดสอบ (สุพัฒน์ สุกมลสันต์, 2538:21)

$$d = \frac{N_u - N_l}{N}$$

เมื่อ d = ค่าอำนาจจำแนก

N_u = จำนวนคนในกลุ่มเก่งจำนวน 27% ที่เลือกตัวเลือก i

N_l = จำนวนคนในกลุ่มอ่อนจำนวน 27% ที่เลือกตัวเลือก i

N = จำนวนคนทั้งหมดในกลุ่ม 27%

3. ความตรงรายข้อของข้อทดสอบ (สุพัฒน์ สุกมลสันต์, 2538:21)

$$r_{pb} = \frac{\bar{x}_r - \bar{x}_w}{S.D._r} \cdot \sqrt{p(1-p)}$$

เมื่อ X_r = ค่าเฉลี่ยของคะแนนผู้สอบที่ตอบข้อทดสอบนั้นถูก

X_w = ค่าเฉลี่ยของคะแนนผู้สอบที่ตอบข้อทดสอบนั้นผิด

SD_r = ค่า SD ของคะแนนรวมของผู้สอบทั้งหมด

p = อัตราส่วนของผู้ที่ตอบข้อทดสอบนั้นถูก

y = ความสูงของแกน Y ในโค้งปกติ ณจุดที่ค่า P

r_{pb} = ค่าอำนาจจำแนกแบบพอยท์ไบเซรียล (Point Biserial Correlation)

4. การทดสอบความมีนัยสำคัญของค่าความตรงรายข้อของข้อทดสอบ (สุพัฒน์ สุกมลสันต์, 2538:21)

$$t = \frac{r_{xz}}{\sqrt{(1-r_{xz}^2)/(n-2)}}$$

เมื่อ r_{xz} = ค่าความเที่ยงแบบ r_{pb} หรือ r_{bis}

n = จำนวนผู้สอบทั้งหมดที่ตอบข้อทดสอบนั้น

5. ความเที่ยงของแบบทดสอบ (สุพัฒน์ สุกมลสันต์, 2538:21)

$$KR_{20} = \frac{k}{k-1} \left[1 - \frac{\sum p_i q_i}{S_i^2} \right]$$

เมื่อ k = จำนวนข้อทดสอบ

p_i = อัตราส่วนของผู้ที่ตอบถูกแต่ละข้อ

q_i = $(1 - p_i)$

S_i^2 = ความแปรปรวนของคะแนนรวมของผู้สอบทั้งหมด

KR_{20} = ค่าความเที่ยงของแบบทดสอบชนิดหนึ่ง

6. ความตรงร่วมสมัย (concurrent validity) ของแบบทดสอบ

ความตรงร่วมสมัยคำนวณจากค่าความสัมพันธ์ระหว่างคะแนนผลการทดสอบของแบบทดสอบ 2 ชุดที่มุ่งทดสอบความรู้ความสามารถอย่างเดียวกัน จากสูตรความสัมพันธ์อย่างง่ายแบบเพียร์สัน (Simple Pearson Product-moment Correlation) ดังนี้ (Downie and Heath, 1970 : 92)

$$r_{xy} = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}}$$

เมื่อ X = คะแนนสอบแบบทดสอบชุดที่ 1 ของผู้สอบคนที่ i
 Y = คะแนนสอบแบบทดสอบชุดที่ 2 ของผู้สอบคนที่ i
 N = จำนวนผู้เข้าสอบทั้งหมด
 r_{xy} = ค่าความตรงร่วมสมัยของแบบทดสอบ

7. ความยากของข้อทดสอบแบบราซ (Rasch model)

การคิดค่าความยากของข้อทดสอบแบบราซ มีค่าโดยประมาณดังนี้ (Lord 1980 : 34)

$$\beta_i \approx \frac{\Delta_i}{r_{bis}}$$

เมื่อ Δ_i = ค่าความยากของข้อทดสอบแบบหนึ่ง
 r_{bis} = ค่าอำนาจจำแนกของข้อทดสอบแบบไบซีเรียล
 β_i = ค่าความยากของข้อทดสอบแบบราซ

8. ความสามารถของผู้สอบ (Wright and Mead, 1977 : 8)

$$\theta_i = H + X \ln \left[\frac{r_v}{(L - r_v)} \right]$$

เมื่อ H = ค่าเฉลี่ยของระดับความยากง่ายทั้งหมด

X = ค่าแก้ไขความคลาดเคลื่อนของ β_i ซึ่งได้จาก

$$X = \sqrt{\frac{1+W^2}{2.89}}$$

เมื่อ W^2 = ความแปรปรวนของค่าความยาก
 L = ความยาวของข้อทดสอบทั้งหมด
 r_v = คะแนนของผู้สอบที่มีระดับความสามารถ r
 \ln = natural log(arithm)
 θ_i = ค่าความสามารถของผู้สอบแต่ละคน

สรุปผลการวิจัย

ผลของการวิจัยตามวัตถุประสงค์ทั้ง 2 ข้อมีรายละเอียดโดยสรุปดังต่อไปนี้

ส่วนที่ 1 การสร้างและพัฒนาแบบทดสอบ CULI-TEPOIC

1. ค่าสถิติโดยสรุปของแบบทดสอบ CULI-TEPOIC แต่ละชุด

ตารางที่ 24: ค่าสถิติโดยสรุปของแบบทดสอบ CULI-TEPOIC แต่ละชุด

ค่าสถิติ	ชุดที่ 1	ชุดที่ 2	ชุดที่ 3
จำนวนข้อทดสอบทั้งหมด	100	100	100
\bar{X} ของคะแนนสอบ	57.871	55.323	53.692
\bar{X} ของค่าความยาก	0.579	0.553	0.537
\bar{X} ของค่าความยากแบบ Δ	12.029	12.376	12.579
\bar{X} ของค่าอำนาจจำแนก	0.334	0.405	0.403
\bar{X} ของค่าความตรงเชิงรายชื่อแบบ r_{pb}	0.383	0.459	0.441
\bar{X} ของค่าความตรงเชิงรายชื่อแบบ r_{bis}	0.292	0.353	0.343
ค่าความเที่ยงแบบ KR_{20}	0.893	0.930	0.926
ค่าความเที่ยงแบบ KR_{21}	0.871	0.914	0.914
ค่าความเที่ยงแบบ Alpha	0.893	0.930	0.926
ค่าความเที่ยงแบบ Split-half	0.865	0.925	0.925

ตารางที่ 24 แสดงให้เห็นว่าแบบทดสอบ CULI-TEPOIC แต่ละชุดมีลักษณะทั่วไปโดยสรุปใกล้เคียงกันมาก เช่น มีค่าความยากโดยเฉลี่ยอยู่ในระดับปานกลางเหมือนกัน ($p = 0.553 - 0.579$) มีค่าอำนาจจำแนกโดยเฉลี่ยอยู่ในระดับที่ดี (DIIA, 2003) และใกล้เคียงกันโดยเฉพาะชุดที่ 2 และ 3 ($d = 0.334$)

-0.405) มีค่าความตรงรายข้อแบบ r_{pb} โดยเฉลี่ยอยู่ในระดับที่ดี (DIIA, 2003) และใกล้เคียงกัน โดยเฉพาะ
ชุดที่ 2 และ 3 ($r_{pb} = 0.292 - 0.353$) และแบบทดสอบ CULI-TEPOIC ชุดที่ 1 มีค่าความเที่ยงอยู่ในระดับ
ดีมาก ส่วนชุดที่ 2 และ 3 มีค่าความเที่ยงอยู่ในระดับดีเยี่ยม (DIIA, 2003) แต่แบบทดสอบทั้ง 3 ชุด
ต่างก็มีค่าดังกล่าวแตกต่างกันไม่มากนัก ($KR_{20} = 0.893 - 0.930$)

2. ค่าความตรงร่วมสมัย (concurrent validity) ของแบบทดสอบทั้ง 3 ชุด และสมการพยากรณ์

ตารางที่ 25: ค่าความตรงร่วมสมัยของแบบทดสอบ CULI-TEPOIC ทั้ง 3 ชุดกับแบบทดสอบ TOEIC

Form	Beta	a (constant)	b	error	n of testees
CULI-TEPOIC 1	0.976	-88.791	12.001	34.990	24
CULI-TEPOIC 2	0.943	-14.464	10.766	52.659	22
CULI-TEPOIC 3	0.904	16.379	10.610	71.537	19

จากตารางที่ 25 แสดงให้เห็นว่าค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนผู้สอบแบบทดสอบ CULI-TEPOIC ชุดที่ 1, 2 และ 3 กับคะแนนสอบแบบทดสอบ TOEIC อยู่ในระดับสูงมากและสูง คือ r_{xy} หรือ Beta = 0.976, 0.943 และ 0.904 ตามลำดับ แสดงว่าแบบทดสอบทั้ง 3 ชุดมีความตรงร่วมสมัยกับแบบทดสอบ TOEIC มาก และจากค่าสถิติต่างในตารางที่ 13 สามารถเขียนเป็นสมการเพื่อพยากรณ์คะแนนสอบ TOEIC จากการทดสอบแบบทดสอบ CULI-TEPOIC ชุดที่ 1, 2 และ 3 ได้ดังนี้

CULI-TEPOIC ชุดที่ 1

$$y' = 12.001 \times \text{คะแนน CULI-TEPOIC 1} + (-88.791)$$

CULI-TEPOIC ชุดที่ 2

$$y' = 10.766 \times \text{คะแนน CULI-TEPOIC 2} + (-14.464)$$

CULI-TEPOIC ชุดที่ 3

$$y' = 10.610 \times \text{คะแนน CULI-TEPOIC 3} + 16.376$$

ดังนั้น จากผลการวิจัยในตารางที่ 24-25 จึงสามารถกล่าวโดยสรุปได้ว่าแบบทดสอบ CULI-TEPOIC ทั้ง 3 ชุดที่สร้างและพัฒนาขึ้นเป็นแบบทดสอบที่มีคุณภาพสูง เช่น มีความยากโดยเฉลี่ยในระดับปานกลาง ($p = 0.537-0.579$) มีความตรงรายข้อในระดับค่อนข้างสูง ($r_{pb} = 0.292-0.353$) มีความเที่ยงในระดับสูง ($KR_{20} = 0.893-0.930$) และมีความตรงร่วมสมัยกับแบบทดสอบที่มีชื่อเสียงคือแบบทดสอบ TOEIC ในระดับสูงและสูงมาก ($r_{xy} = 0.904-0.976$) เป็นต้น

ส่วนที่ 2 การสร้างและพัฒนาโปรแกรมคอมพิวเตอร์ที่ใช้ในการพัฒนาแบบทดสอบและการทดสอบ

1. ดำเนินการจัดหาและจัดจ้างผู้มีความชำนาญในการเขียน โปรแกรมคอมพิวเตอร์ด้วยภาษา Visual BASIC เพื่อใช้กับระบบ Windows และมีความเข้าใจภาษา Quick BASIC, FORTRAN และ Fox Pro เป็นอย่างดี เนื่องจากโปรแกรมต่างๆที่ต้องการพัฒนานั้นเดิมผู้วิจัยเขียนด้วยภาษาดังกล่าวแต่ใช้งานได้เฉพาะในระบบ DOS (Disk Operation System) เท่านั้น

2. อธิบายให้ผู้เขียนโปรแกรมเข้าใจการทำงานของทั้ง 5 โปรแกรมที่ผู้วิจัยต้องการพัฒนาให้ใช้ได้กับระบบ Windows รวมทั้งความต้องการให้ผู้เขียนโปรแกรมเขียนด้วยภาษา Visual BASIC เพราะว่าโครงสร้างของภาษาดังกล่าวคล้ายคลึงกันกับของภาษา Quick BASIC และ FORTRAN จะทำให้ผู้เขียนโปรแกรมทำงานได้รวดเร็วและถูกต้องยิ่งขึ้น ส่วนระบบฐานข้อมูลให้ใช้ภาษา Microsoft Access แทนภาษา Fox Pro อนึ่ง โปรแกรมทั้ง 5 โปรแกรมได้แก่
 1. โปรแกรมวิเคราะห์ข้อทดสอบแบบประเพณีนิยม (Classical Model)
 2. โปรแกรมวิเคราะห์ข้อทดสอบแบบราช (Rasch Model)
 3. โปรแกรมธนาคารข้อทดสอบ (Item Banking)
 4. โปรแกรมการทดสอบแบบปรับเปลี่ยน (Adaptive Testing) และ
 5. โปรแกรมกำหนดระดับคะแนน (Grade Assignment)
3. มอบรหัสต้นฉบับ (source codes) ทั้งที่เป็นสำเนาอิเล็กทรอนิกส์ (soft copy) และสำเนาแสดงผล (hard copy) ให้กับผู้เขียนโปรแกรม เพื่อให้ดำเนินการพัฒนาตามที่ได้ตกลงกันไว้
4. ตรวจสอบการทำงานของโปรแกรมที่พัฒนาขึ้นเป็นระยะๆ และให้คำแนะนำแก่ผู้เขียนโปรแกรมเป็นระยะๆ เพื่อให้ผลงานเป็นไปตามที่ต้องการและถูกต้องมากที่สุด
5. เมื่อโปรแกรมเสร็จสมบูรณ์แล้ว ผู้วิจัยทำการตรวจสอบการทำงานของโปรแกรมใหม่ทั้ง 5 โปรแกรม โดยการเปรียบเทียบกับผลการทำงานของโปรแกรมเดิมว่าเหมือนกันหรือไม่

ผลของการวิจัยตามวัตถุประสงค์ข้อที่ 2 มีดังต่อไปนี้

1. โปรแกรมที่สร้างและพัฒนาขึ้นทั้ง 5 โปรแกรมสามารถทำงานได้ถูกต้อง และครบถ้วนเหมือนเช่นเดียวกับโปรแกรมเดิม แต่ใช้งานได้สะดวกกว่าโปรแกรมชุดเดิมเนื่องจากทั้ง 5 โปรแกรมทำการติดตั้งแยกกัน 2 ครั้งและทำงานร่วมกันได้แทนที่จะเป็น 4 ครั้งเหมือนโปรแกรมชุดเดิม ซึ่งมีบางโปรแกรมทำงานเป็นอิสระ
2. โปรแกรมที่สร้างและพัฒนาขึ้นหนึ่ง โปรแกรมสามารถทำงานได้ก้าวหน้ากว่าโปรแกรมเดิมคือโปรแกรมธนาคารข้อทดสอบ กล่าวคือสามารถจัดเก็บภาพประกอบ (graphics) ตรวจสอบรหัสความปลอดภัยของฐานข้อมูล และจัดการฐานข้อมูลตามอภิธานุกรมได้ซับซ้อนมากกว่าโปรแกรมเดิม

ตัวอย่างผลการวิจัยของส่วนที่ 2 นี้ คือผลการเปรียบเทียบการทำงานของโปรแกรม Classical กับโปรแกรม CTIA 7.0 และโปรแกรม Item Bank กับโปรแกรม ITB Pro ดังต่อไปนี้

1. การเปรียบเทียบผลการทำงานของโปรแกรม Classical กับโปรแกรม CTIA 7.0

ตารางที่ 26: การเปรียบเทียบผลการทำงานของโปรแกรม Classical กับโปรแกรม CTIA 7.0 จากการวิเคราะห์แบบทดสอบชุดเดียวกัน

ค่าสถิติ *	Classical	CTIA 7.0
1. จำนวนข้อทดสอบทั้งหมด	100	100
2. \bar{X} ของคะแนนสอบ	57.871	57.871
3. \bar{X} ของค่าความยาก	0.579	0.579
4. \bar{X} ของค่าความยากแบบ Δ	12.029	12.029
5. \bar{X} ของค่าอำนาจจำแนก	0.334	0.334
6. \bar{X} ของค่าความตรงรายข้อแบบ r_{bis}	0.383	0.383
7. \bar{X} ของค่าความตรงรายข้อแบบ r_{pb}	0.292	0.292
8. ค่าความเที่ยงแบบ KR_{20}	0.893	0.893
9. ค่าความเที่ยงแบบ KR_{21}	0.871	0.871
10. ค่าความเที่ยงแบบ Alpha	0.893	0.893
11. ค่าความเที่ยงแบบ Split-half	0.865	0.865
12. จำนวนข้อทดสอบที่ควรปรับปรุง	15	15

* $r_{xy} = 1.00$ ($p < 0.05$)

ผลจากการวิเคราะห์แบบทดสอบชุดเดียวกันในตารางที่ 26 ปรากฏว่าทุกค่าในมิติต่างๆ เช่น ค่าความยากเฉลี่ย ค่าอำนาจจำแนกเฉลี่ย ค่าความตรงรายข้อเฉลี่ย และค่าความเที่ยงแบบต่างๆ มีขนาดเท่ากันพอดี และเมื่อทำการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าสถิติต่างๆ ที่ได้จากโปรแกรมทั้งสอง ปรากฏว่า $r_{xy} = 1.00$ ($p < 0.05$) หมายความว่าทุกค่ามีขนาดเท่ากันพอดี ซึ่งแสดงว่าโปรแกรม Classical ทำการคำนวณได้ค่าสถิติต่างๆ โดยสรุปจากการวิเคราะห์ข้อทดสอบรายข้อแบบประเพณีนิยม ได้ถูกต้องเช่นเดียวกับโปรแกรม CTIA 7.0

2. ความสามารถในการทำงานของโปรแกรม Item Bank กับโปรแกรม ITB Pro

ตารางที่ 27: การเปรียบเทียบความสามารถในการทำงานของโปรแกรม Item Bank กับโปรแกรม ITB Pro

ความสามารถในการทำงาน	Item Bank	ITB Pro
1. การสร้างข้อทดสอบ	ทำได้ถูกต้อง	ทำได้
2. การแก้ไขข้อทดสอบ	ทำได้ถูกต้อง	ทำได้
3. การลบข้อทดสอบทิ้ง	ทำได้ถูกต้อง	ทำได้
4. การเพิ่มข้อทดสอบ	ทำได้ถูกต้อง	ทำได้
5. การเปลี่ยนรหัสข้อทดสอบ	ทำได้ถูกต้อง	ทำได้
6. การเก็บภาพประกอบ	ทำได้ถูกต้อง	ทำไม่ได้
7. การสุ่มเลือกข้อทดสอบ	ทำได้	ทำได้
8. การแสดงข้อทดสอบที่เลือกแล้ว	ทำได้ถูกต้อง	ทำได้
9. การจัดพิมพ์ข้อทดสอบ	ทำได้ถูกต้อง	ทำได้
10. การจัดรูปแบบของแบบทดสอบ	ทำได้	ทำได้
11. การสำรองฐานข้อมูล	ทำได้ถูกต้อง	ทำได้
12. การตรวจรหัสความปลอดภัยของฐานข้อมูล	ทำได้ถูกต้อง	ทำไม่ได้
13. การจัดการฐานข้อมูลตามอภิธานุกรม	ทำได้ซับซ้อนกว่า	ทำได้
14. การใช้กับภาษาไทยและภาษาอังกฤษ	ทำได้ถูกต้อง	ทำได้
15. การจัดการต่างๆดังกล่าวกับแบบทดสอบ	ทำได้ถูกต้อง	ทำได้

จากการเปรียบเทียบความสามารถในการทำงานของโปรแกรม Item Bank กับโปรแกรม ITB Pro จากฐานข้อมูลชุดเดียวกัน ปรากฏว่าโปรแกรมใหม่ที่พัฒนาขึ้นสามารถทำงานต่างๆที่เกี่ยวข้องกับการทำงานของธนาคารข้อทดสอบได้ดีเช่นเดียวกับโปรแกรมเดิม เช่น การเตรียมข้อทดสอบและแบบทดสอบชนิดต่างๆเข้าเก็บไว้ในฐานข้อมูล การปรับปรุงแก้ไข การคัดเลือกข้อทดสอบหรือแบบทดสอบ และการจัดพิมพ์สิ่งที่ได้เลือกแล้วในรูปแบบที่ต้องการได้ นอกจากนี้ยังพบว่าโปรแกรม Item Bank ยังสามารถทำงาน 3 อย่างได้ดีกว่าโปรแกรม ITB Pro คือ สามารถตรวจรหัสความปลอดภัยของฐานข้อมูลได้ และจัดเก็บภาพประกอบข้อทดสอบหรือแบบทดสอบได้ ซึ่งความสามารถดังกล่าวนี้ไม่มีในโปรแกรม ITB Pro รวมทั้งสามารถจัดการเก็บและเลือกข้อมูลจากฐานข้อมูลที่แยกประเภทตามอนุกรมวิธาน (taxonomy) ได้ซับซ้อนมากกว่า

อภิปรายผล

ผลของการวิจัยครั้งนี้มีประเด็นที่น่าสนใจดังต่อไปนี้

1. ทำไมแบบทดสอบ CULI-TEPOIC ที่สร้างขึ้นแต่ละชุดจึงมีคุณภาพสูงและมีความตรงร่วมสมัยกับแบบทดสอบ TOEIC สูงมากทั้งที่มีจำนวนข้อทดสอบน้อยกว่า

การที่ผลการวิจัยเป็นเช่นนี้อาจเป็นเพราะเหตุผลต่อไปนี้ คือ

1. คณะกรรมการสร้างแบบทดสอบเป็นผู้ที่มีความสามารถและประสบการณ์ในการสร้างแบบทดสอบมาเป็นอย่างดี รวมทั้งมีความรู้และความสามารถทางภาษาอังกฤษในระดับที่ดี ทั้งนี้เพราะว่าเป็นอาจารย์สอนภาษาอังกฤษในสถาบันภาษาของจุฬาลงกรณ์มหาวิทยาลัย ซึ่งเป็นสถาบันที่มีชื่อเสียงและเป็นที่ยอมรับกันทั่วไป ดังนั้นผลงานในการสร้างแบบทดสอบจึงมีคุณภาพสูง โดยเฉพาะในด้านความตรงรายข้อ (item validity) และความเที่ยงของแบบทดสอบทั้งฉบับ ดังปรากฏในผลการวิจัยครั้งนี้ กล่าวคือแบบทดสอบ CULI-TEPOIC ชุดที่ 1, 2 และ 3 มีความตรงรายข้อโดยเฉลี่ย = 0.383, 0.459 และ 0.441 ตามลำดับ ซึ่งสูงกว่าเกณฑ์ทั่วไปคือ 0.25 (Cheung and Bucat, 2002; DIIA, 2003; Runte, 2001; Jerard, 1995; สุพัฒน์ สุขมลสันต์, 2542; Ebel, 1972) และมีความเที่ยงแบบ $KR_{20} = 0.893, 0.930$ และ 0.926 ซึ่งขนาดสูงในระดับดีมาก และคิยอคเยี่ยม เนื่องจากสูงกว่าเกณฑ์ทั่วไปคือ $KR_{20} = > 0.80$ (Jerard, 1995)
2. รูปแบบ (format) ของแบบทดสอบ CULI-TEPOIC มีลักษณะคล้ายกับแบบทดสอบ TOEIC มาก กล่าวคือแบบทดสอบทั้งสองชนิดเป็นแบบทดสอบปรนัยชนิดเลือกตอบที่มีตัวเลือกจำนวน 4 ตัวเลือกเพื่อทดสอบความรู้และความสามารถทั่วไปทางด้านการฟังเข้าใจความ และการเขียนและการอ่านเข้าใจความเหมือน แต่ต่างกันที่จำนวนข้อทดสอบ เพราะว่าแบบทดสอบ CULI-TEPOIC มีข้อทดสอบจำนวนส่วนละ 50 ข้อรวมเป็นข้อทดสอบทั้งหมด 100 ข้อ แต่ว่าแบบทดสอบ TOEIC มีข้อทดสอบส่วนละ 100 ข้อรวมเป็นข้อทดสอบทั้งหมด 200 ข้อ ทำให้แบบทดสอบทั้งสองทั้งสองชนิดใช้เวลาสอบแตกต่างกันด้วย นอกจากนี้แบบทดสอบทั้งสองชนิดยังมุ่งทดสอบความรู้และความสามารถของผู้สอบที่คล้ายคลึงกันมาก กล่าวคือแบบทดสอบ TOEIC มีเนื้อหาที่มุ่งทดสอบความรู้และความสามารถทั่วไปในการใช้ภาษาอังกฤษเพื่อการสื่อสารนานาชาติในบริบทต่างๆในชีวิตประจำวัน ส่วนแบบทดสอบ CULI-TEPOIC มีเนื้อหาที่มุ่งทดสอบความรู้และความสามารถทั่วไปในการใช้ภาษาอังกฤษเพื่อการสื่อสารนานาชาติในบริบทต่างๆของวิชาชีพ
3. ผู้สอบแบบทดสอบ CULI-TEPOIC แม้วามีจำนวนไม่มากนักแต่มีความรู้และความสามารถทางภาษาอังกฤษหลากหลายระดับ (heterogeneous group) ซึ่งอาจสังเกตได้จากพิสัย (range) ของคะแนนสอบและดัชนีการกระจายของคะแนนสอบ (coefficient of variation) กล่าวคือแบบทดสอบชุดที่ 1, 2 และ 3 มีพิสัยของคะแนน = 65, 67 และ 71 และดัชนีการกระจายของ

คะแนน = 22.98%, 29.19% และ 30.17% ตามลำดับ (ดูตารางที่ 11) นอกจากนี้เมื่อสังเกตดูค่าความเบ้ (skewness) ของแบบทดสอบทั้ง 3 ชุดมีค่า = -0.02, 0.09 และ 0.19 ซึ่งว่ามีค่าใกล้เคียงกับ 0.00 แสดงว่าประมาณร้อยละ 50 ของผู้สอบแบบทดสอบแต่ละชุดมีคะแนนจากการทดสอบมากกว่าหรือน้อยกว่าคะแนนค่าเฉลี่ยของแบบทดสอบ ค่าต่างๆดังกล่าวแล้วข้างต้นแสดงให้เห็นว่าความรู้และความสามารถของผู้สอบแบบทดสอบแต่ละชุดมีความหลากหลายและอาจเป็นปัจจัยหนึ่งที่ทำให้ค่าทางสถิติของแบบทดสอบแต่ละชุดเป็นค่าที่สูงและน่าเชื่อถือ (Jerard, 1995; Cronbach, 1960: 133) เพราะว่าการกลุ่มผู้สอบที่มีความสามารถหลากหลายเป็นลักษณะของตัวแทนที่ดีของประชากรของผู้เข้ารับการทดสอบแม้ว่าจำนวนจะไม่มากนัก

4. ผู้สอบแบบทดสอบ TOEIC มีความสามารถหลากหลายเนื่องจากได้รับการสุ่มอย่างมีระบบ (systematic random sampling technique) จากผู้วิจัยโดยอาศัยคะแนนสอบจากแบบทดสอบ CULI-TEPOIC ที่เรียงลำดับจากคะแนนสูงสุดไปหาคะแนนต่ำสุดเป็นเกณฑ์ ดังนั้นคะแนนสอบที่ได้จากการสอบแบบทดสอบทั้ง 2 ชุดจึงมีค่าสัมประสิทธิ์สหสัมพันธ์กันสูงมาก กล่าวคือ มีค่า r_{xy} ของแบบทดสอบชุดที่ 1, 2 และ 3 = 0.976, 0.943 และ 0.904 ตามลำดับ

2. การที่แบบทดสอบ CULI-TEPOIC แต่ละชุดมีคุณภาพสูงและมีความตรงร่วมสมัยกับแบบทดสอบ TOEIC สูงมาก ผู้สอบสามารถใช้ผลการสอบของแบบทดสอบ CULI-TEPOIC แทนผลการสอบของแบบทดสอบ TOEIC ได้หรือไม่

คำตอบของประเด็นข้างต้นคือ “ไม่ได้โดยตรง” แต่ “สามารถเทียบเคียงได้” เพราะเหตุผลดังนี้คือ

1. แบบทดสอบ CULI-TEPOIC และ TOEIC ต่างก็เป็นแบบทดสอบวัดสมรรถภาพทั่วไปในการใช้ภาษาอังกฤษเพื่อการสื่อสารนานาชาติ แต่แบบทดสอบ CULI-TEPOIC เน้นที่การสื่อสารนานาชาติในบริบททั่วไปในวิชาชีพ ส่วนแบบทดสอบ TOEIC เน้นที่การสื่อสารนานาชาติในบริบททั่วไปในชีวิตประจำวัน ดังนั้นหน้าที่ของภาษา (language functions) ในข้อทดสอบแม้ว่าจะมีบางส่วนเหมือนกัน แต่ก็ไม่เหมือนกันทั้งหมด
2. แบบทดสอบ CULI-TEPOIC และ TOEIC ไม่ได้เป็นแบบทดสอบคู่ขนานจริง (true parallel forms) กล่าวคือ ไม่ได้สร้างจากตารางกำหนดลักษณะเฉพาะ (test specification table) เดียวกัน จึงมีเนื้อหาต่างกัน จำนวนข้อทดสอบแตกต่างกัน มีระดับความยากต่างกัน และการกระจายของคะแนนต่างกัน กล่าวคือ แบบทดสอบ CULI-TEPOIC ชุดที่ 1, 2 และ 3 มีคะแนนเฉลี่ย = 57.871, 55.323 และ 53.692 และค่า S.D. = 13.30, 16.15 และ 16.20 ตามลำดับ (ดูตารางที่ 11) ส่วนแบบทดสอบ TOEIC มีคะแนนเฉลี่ยของผู้สอบจำนวน 65 คน = 124.318 (จากคะแนนเต็ม = 200) และ S.D. = 39.28 หรือ = 62.153 และ และ 19.64 เมื่อเทียบเป็นคะแนนเต็มเท่ากับ 100 คะแนนเท่ากัน

3. ตารางเปรียบเทียบคะแนนเกิดจากสมการพยากรณ์ทางสถิติที่ค่าสัมประสิทธิ์สหสัมพันธ์ หรือค่า Beta (β) ยังไม่สมบูรณ์ กล่าวคือยังมีค่าน้อยกว่า 1.0 แม้ว่าจะมีค่าสูงมาก คือ 0.904 – 0.976 จึงยังมีค่าความคลาดเคลื่อนในการพยากรณ์ (คุณภาพผนวก ค. และตารางที่ 25)
3. ทำไมโปรแกรมที่สร้างและพัฒนาขึ้นทั้ง 5 โปรแกรมจึงสามารถทำงานได้ถูกต้องและครบถ้วน เหมือนกับโปรแกรมเดิม และมีบางโปรแกรมทำงานได้มากกว่าและดีกว่าโปรแกรมเดิม
 - การที่ผลการวิจัยเป็นเช่นนี้อาจเป็นเพราะเหตุผลต่อไปนี้ คือ
 1. โปรแกรมใหม่ทั้ง 5 โปรแกรมสร้างจากรหัสต้นฉบับ (source codes) ทั้งที่เป็นสำเนาอิเล็กทรอนิกส์ (soft copy) และสำเนาแสดงผล (hard copy) ทำให้ผู้เขียนโปรแกรมที่มีความรู้ความสามารถในภาษาเดิมคือ Quick BASIC, FORTRAN และ Fox Pro และภาษาใหม่คือ Visual BASIC และ Microsoft Access 2000 สามารถปรับเปลี่ยนหรือถ่ายโอน (transfer) ภาษาของโปรแกรมเดิมให้เป็นภาษาใหม่ได้ไม่ยากนัก ดังนั้นรูปลักษณะของผลลัพธ์ (output) ของโปรแกรมเดิมและโปรแกรมใหม่จึงเหมือนกัน
 2. รหัสต้นฉบับ (source codes) ของแต่ละโปรแกรมมีขั้นตอนวิธีการทำงาน (algorithm) แฝงอยู่ ดังนั้นโปรแกรมใหม่ที่สร้างขึ้นจึงมีกระบวนการในการทำงานเหมือนเดิมทุกประการ ดังนั้นการทำงานของโปรแกรมใหม่ย่อมถูกต้องและครบถ้วนเหมือนกับของโปรแกรมเดิม
 3. การที่มีบางโปรแกรม คือ โปรแกรมธนาคารข้อทดสอบ สามารถทำงานบางอย่างได้มากกว่าและดีกว่าโปรแกรมเดิมนั้น เป็นเพราะว่าภาษาคอมพิวเตอร์และเครื่องคอมพิวเตอร์ในปัจจุบันนี้มีความก้าวหน้ามากกว่าตอนที่เขียนโปรแกรมเดิมมาก กล่าวคือ ปัจจุบันนี้เครื่องคอมพิวเตอร์สามารถจัดเก็บไฟล์รูปภาพ ไฟล์เสียง และไฟล์ภาพเคลื่อนไหวได้และภาษาคอมพิวเตอร์ที่ใช้เขียนโปรแกรมบางภาษาสามารถจัดการกับไฟล์ต่างๆดังกล่าวได้ ด้วยเหตุนี้ โปรแกรมธนาคารข้อทดสอบใหม่จึงเขียนขึ้นเพื่อจัดเก็บไฟล์รูปภาพด้วยนอกเหนือจากไฟล์ตัวอักษรแล้ว ส่วนไฟล์เสียง และไฟล์ภาพเคลื่อนไหวนั้นยังไม่ได้จัดเก็บเนื่องจากมีปัญหาในการสุ่มเลือกเมื่อต้องการสร้างแบบทดสอบชุดใหม่

ข้อเสนอแนะ

สืบเนื่องจากผลการวิจัยดังกล่าวแล้วข้างต้นทั้ง 2 ข้อ ผู้วิจัยมีข้อเสนอแนะเพื่อนำผลการวิจัยไปใช้ และเพื่อการวิจัยต่อไปในอนาคต ดังต่อไปนี้

ก. เพื่อนำผลการวิจัยไปใช้

1. เนื่องจากค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนสอบของแบบทดสอบ CULI-TEPOIC ชุดที่ 1, 2 และ 3 กับคะแนนสอบของแบบทดสอบ TOEIC ยังไม่สมบูรณ์ดังได้กล่าวมาแล้ว ทำให้การพยากรณ์คะแนนผลการทดสอบของแบบทดสอบ TOEIC จากคะแนนผลการสอบของ

แบบทดสอบ CULI-TEPOIC ยังมีค่าความคลาดเคลื่อน ดังนั้น ทุกครั้งที่หน่วยงานที่รับผิดชอบในการเปิดเผยค่าคะแนนเปรียบเทียบผลการสอบควรแจ้งให้ผู้รับบริการทราบค่าความคลาดเคลื่อนดังกล่าวด้วยทุกครั้ง รวมทั้งในเอกสารทางราชการที่ระบุค่าคะแนนเปรียบเทียบดังกล่าวด้วย

2. เนื่องจากแบบทดสอบ CULI-TEPOIC และ TOEIC ไม่ได้เป็นแบบทดสอบคู่ขนานจริง (true parallel forms) ดังได้กล่าวมาแล้ว ดังนั้น ผู้ให้บริการการทดสอบแบบทดสอบ CULI-TEPOIC ควรทำความเข้าใจกับผู้ขอรับบริการให้ถูกต้องว่า คะแนนเปรียบเทียบระหว่างผลการสอบจากแบบทดสอบ CULI-TEPOIC กับ TOEIC เป็นเพียงคะแนนพยากรณ์ทางสถิติที่คาดว่าผู้สอบจะสามารถสอบได้เท่านั้น ไม่ใช่คะแนนที่เป็นจริง และคะแนนจากแบบทดสอบ CULI-TEPOIC ไม่สามารถใช้แทนคะแนนของแบบทดสอบ TOEIC ได้ทุกกรณี กล่าวคือ คะแนน “ใช้แทนกันไม่ได้โดยตรง” แต่ว่า “สามารถเทียบเคียงกันได้”
3. เนื่องจากโปรแกรมที่พัฒนาขึ้นบางโปรแกรมอาศัยแนวคิดจากทฤษฎีการตอบสนองของข้อทดสอบ (Item Response Theory) เช่น โปรแกรมการวิเคราะห์ข้อทดสอบแบบราช (Rasch Model) และโปรแกรมทดสอบแบบปรับเปลี่ยน ดังนั้น ผู้ที่จะใช้โปรแกรมดังกล่าวควรมีพื้นฐานความรู้ทางด้านการวัดและประเมินผลทางการศึกษาที่เกี่ยวข้องการแนวคิดของทฤษฎีดังกล่าวด้วยจึงจะสามารถเข้าใจผลลัพธ์ (output) จากการวิเคราะห์ข้อสอบจากโปรแกรมดังกล่าวได้
4. หน่วยงานใดที่ให้บริการการประเมินผลสัมฤทธิ์ในการเรียน เช่น งานทะเบียนของสถานศึกษา หรือหน่วยงานที่ให้บริการการพัฒนาแบบทดสอบให้เป็นมาตรฐาน รวมทั้งให้บริการการสอบแก่บุคคลอื่น เช่น ศูนย์บริการการสอบ และสถาบันการทดสอบต่างๆ ควรจะนำโปรแกรมที่ได้จากการวิจัยครั้งนี้ไปใช้ในหน่วยงานดังกล่าว เพื่อให้งานประจำสามารถดำเนินไปได้อย่างถูกต้อง รวดเร็ว และน่าเชื่อถือได้มากยิ่งขึ้น และที่สำคัญมากก็คือ เพื่อให้งานการทดสอบ วัดและประเมินผลเป็นงานเชิงวิทยาศาสตร์ที่สามารถตรวจสอบได้ทุกขั้นตอน

ข. เพื่อการวิจัยต่อไปในอนาคต

1. เนื่องจากขณะนี้แบบทดสอบ CULI-TEPOIC ทั้ง 3 ชุดอยู่ในความรับผิดชอบของศูนย์ทดสอบของสถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย ดังนั้นศูนย์ทดสอบควรดำเนินการวิจัยต่อเนื่องเพื่อพัฒนาแบบทดสอบทั้ง 3 ชุดให้มีคุณภาพดียิ่งขึ้น โดยเฉพาะการปรับแก้ข้อทดสอบจำนวนหนึ่งที่มีคุณภาพยังไม่ดี และคัดเลือกจำนวนหนึ่งของข้อทดสอบที่อยู่ในเกณฑ์ที่เป็นข้อทดสอบที่ดี และถ้าเป็นไปได้ควรทำการวิจัยเพื่อหาค่าความตรงร่วมสมัยกับแบบทดสอบมาตรฐานอีกภายหลังการปรับปรุงแก้ไขข้อทดสอบต่างๆ แล้ว
2. ศูนย์ทดสอบควรรับผิดชอบในการทำการวิจัยเพื่อสร้างและพัฒนาแบบทดสอบ CULI-TEPOIC และแบบทดสอบอื่นๆ อีกตามความต้องการของสังคมและผู้ขอรับบริการ โดยแบบทดสอบแต่ละชุดก่อนที่จะให้บริการควรจะได้ผ่านการวิเคราะห์ข้อทดสอบก่อนอย่างน้อยที่สุดในระดับ

ความตรงเชิงเนื้อหา ทั้งนี้เพื่อเป็นการประกันคุณภาพขั้นต่ำของแบบทดสอบ และหากทำได้ก็ควรทำการวิเคราะห์ข้อทดสอบเพื่อหาคุณภาพด้านอื่นๆที่จำเป็นต่อไป

3. ผู้สนใจควรทำการวิจัยเพื่อสร้าง โปรแกรมการทดสอบแบบปรับเปลี่ยนที่ชี้คเนื้อหาเป็นเกณฑ์ (Computerized Content-based Adaptive Testing Program) และโปรแกรมการทดสอบแบบปรับเปลี่ยนตามแนวทฤษฎีการตอบสนองต่อข้อทดสอบ เช่นอาศัยแนวคิดของรูปแบบ 1 พารามิเตอร์ (One-Parameter Model) รูปแบบ 2 พารามิเตอร์ (Two-Parameter Model) และรูปแบบ 3 พารามิเตอร์ (Three-Parameter Model) เนื่องจากแนวโน้มของการทดสอบในอนาคตจะมีการทดสอบแบบปรับเปลี่ยนมากขึ้น
4. ผู้สนใจควรทำการวิจัยเพื่อพัฒนาโปรแกรมธนาคารข้อทดสอบและแบบทดสอบด้วยเครื่องคอมพิวเตอร์ที่สามารถจัดเก็บไฟล์ต่างๆ นอกเหนือจากไฟล์ตัวอักษรและไฟล์รูปภาพ เช่น ไฟล์เสียง ไฟล์ภาพเคลื่อนไหว (animation) ไฟล์ภาพยนตร์ และไฟล์วีดิทัศน์ เป็นต้น เนื่องจากแนวโน้มของการทดสอบในอนาคตจะเป็นการทดสอบตามสภาพจริง (authentic assessment) ดังนั้นไฟล์ต่างๆเหล่านี้จึงมีความจำเป็นต่อการทดสอบ

สุวัฒน์ สุขมลสันต์

7 สิงหาคม 2549

บรรณานุกรม

- บึงอร สว่างวโรรส. (2527). "การศึกษาความสอดคล้องระหว่างภาษาอังกฤษธุรกิจที่สอนในมหาวิทยาลัยในประเทศไทยกับความต้องการของวงการธุรกิจ." รายงานการวิจัย. กรุงเทพมหานคร: สถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย.
- ศิริชัย กาญจนวาสี. (2545). ทฤษฎีการทดสอบแนวใหม่. กรุงเทพฯ: คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.
- สถาบันภาษา. (2542). รายงานประจำปี 2542 . กรุงเทพมหานคร: สถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย.
- สุพัฒน์ สุกมลสันต์ และคณะ (2529). "การประเมินผลหลักสูตรภาษาอังกฤษพื้นฐาน." รายงานการวิจัย. กรุงเทพมหานคร: สถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย
- สุพัฒน์ สุกมลสันต์. (2536). "การสอบสมิทธิภาพทั่วไปทางภาษาอังกฤษครั้งที่ 3 ปี พ.ศ. 2535 ในโครงการสร้างและพัฒนาคุณภาพแบบทดสอบสมิทธิภาพทั่วไปทางภาษาอังกฤษและระบบการทดสอบที่เกี่ยวข้อง". รายงานการวิจัย. กรุงเทพมหานคร: สถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย.
- สุพัฒน์ สุกมลสันต์. (2538). การวิเคราะห์ข้อทดสอบแนวใหม่ด้วยคอมพิวเตอร์. กรุงเทพฯ: บริษัทวิทย์พัฒน์ จำกัด
- สุพัฒน์ สุกมลสันต์. (2539). ธนาคารข้อทดสอบและการทดสอบปรับเปลี่ยนด้วยคอมพิวเตอร์. กรุงเทพฯ: บริษัทวิทย์พัฒน์ จำกัด
- สุพัฒน์ สุกมลสันต์. (2542). การวิเคราะห์ข้อทดสอบและตัดเกรดด้วยคอมพิวเตอร์. กรุงเทพฯ: บริษัทวิทย์พัฒน์ จำกัด
- สุพัฒน์ สุกมลสันต์ และคณะ. (2533). "แนวโน้มนระดับสมิทธิภาพทั่วไปทางภาษาอังกฤษของนิสิตจุฬาลงกรณ์มหาวิทยาลัย และบุคคลภายนอกที่สนใจปี พ.ศ. 2533," รายงานการวิจัย. กรุงเทพฯ: สถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย
- อัจฉรา วงศ์โสธร และคณะ.(2523). "การสำรวจความต้องการของสังคมต่อการใช้ภาษาอังกฤษ." รายงานการวิจัย. กรุงเทพมหานคร: สถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย
- อัจฉรา วงศ์โสธร และคณะ.(2535). "การสอบสมิทธิภาพทั่วไปทางภาษาอังกฤษครั้งที่ 2 ปี พ.ศ. 2534 ในโครงการสร้างและพัฒนาคุณภาพแบบทดสอบสมิทธิภาพทั่วไปทางภาษาอังกฤษและระบบการทดสอบที่เกี่ยวข้อง." รายงานการวิจัย. กรุงเทพมหานคร: สถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย.
- อุทัย บุญประเสริฐ. (2535). การจัดทำคลังข้อทดสอบ. กรุงเทพฯ: ภาควิชาบริหารการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

- Alderson, J. C. (1988). *New procedures for validating proficiency tests of ESP? Theory and practice.* *Language Testing*, 5, 220-32.
- Alderson, J. C. (2000). *Assessing reading.* Cambridge, UK: Cambridge University Press.
- Alderson, J. C. and Others. (1995). *Language Test Construction and Evaluation.*
- Allen, J.P.B. and Davies, A.(1977). *Testing and Experimental Methods*, (The Edinburgh Course in Applied Linguistics, Volume 4.) London : Oxford University Press.
- Anastasi, A.(1965). *Psychological Testing*, New York: The Macmillan Co.
- Anderson, R. C. (1972). "How to construct achievement tests to assess comprehension." *Review of Educational Research* 42: 145-170.
- Anderson, R. C. (1981). *New Dimensions in Second Language Research.* Rowley, Mass: Newbury House.
- Bachman, L. F. (1990). *Fundamental considerations in language testing.* Oxford: Oxford University Press.
- Bachman, L. F. and Palmer, A. S. (1996). *Language testing in practice.* Oxford: Oxford University Press.
- Baker, D. (1989). *Language Testing: A Critical Survey and Practical Guide.* Oxford: Oxford University Press.
- Baker, F.B. (2001). *The Basic of Item Response Theory.* ERIC Clearing House on Assessment and Evaluation.
- Bloom, B. and Others (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain.* New York, Toronto: Longmans, Green.
- Bloom, B.S. (ed.) (1968). *Taxonomy of Educational Objectives: Handbook I: Cognitive Domain.* New York: David McKay
- Brennan, R. L. (1992). *Elements of Generalizability Theory.* New York: American College Testing.
- Brown, A. and Others (1994). "Mapping Abilities and Skill Levels Using Rasch Techniques." *Occasional Papers in Language Testing.* Melbourne University Press.
- Brown, J. D. (2000). *What is Construct Validity? JALT Testing and Evaluation SIG Newsletter*, 4, 2, 7-10.
- Brown, J. D. (2001). "Pragmatic tests: Different purposes, different tests." In G. Kasper and K. R. Rose (Eds.), *Pragmatics in language teaching* (pp. 301-325). Cambridge: Cambridge University Press
- Burton, S.J. and Others. (1991). *How to Write Better Multiple-Choice Items.* Brigham Young University Testing Service. Retrieved from <http://testing.byu.edu/info/handbooks/betteritems.pdf> on June 2002.
- Canale, M. and Swain, M. (1980). "Theoretical bases of communicative approaches to second-language teaching and testing." *Applied Linguistics*, 1 (1), 1-47.
- Canale, M. (1983). "On Some Dimensions of Language Proficiency." In J. Oller (ed.) *Issues in Language Testing Research.* Rowley, Mass.: Newbury House. 333-342.
- Canale, M. and Swain, M. (1980). "Theoretical bases of communicative approaches to second language teaching and testing." *Applied Linguistics* 1:1-47.
- Carroll, J.B. (1972). "Foreign Language Testing: Will the Persistent Problems Persist?" in M. C. O'Brien (ed.), *Testing in Second Language Teaching*, Ireland: ATESOL.
- Cheung, D. and Bucat, R. (2002). *How can we construct good Multiple-Choice Items.* Retrieved from <http://www3.fed.cuhk.edu.hk/chemistry/files/constructMC.pdf> on July 12, 2004.
- Clark, J.L.D. (1972). *Foreign Language Testing: Theory and Practice.* PA: The Center for Curriculum Development.

- Clegg, V. and Cashin, W.E. (1986). *Improving Multiple-Choice Tests*. Retrieved from http://www.idea.ksu.edu/papers/Idea_Paper_16.pdf on July 12, 2004.
- Clymer, T. (1972). "What is reading?: Some Current Concepts." In A. Melnik and J. Merritt (eds.) *Reading Today and Tomorrow*. London: University of London Press Limited.
- Cronbach, L.J. (1960). *Essentials of Psychological Testing*. New York: Harper and Brothers.
- Davies, A. (1972). "Language Proficiency Testing and the Syllabus." In M.C. O'Brien (ed.), *Testing in Second Language Teaching : New Dimensions*. Ireland, ATESOL.
- Davies, A. (1979). "Language Testing Survey Articles I & II" in *Language Teaching and Linguistics Abstracts*.
- Davies, A. (ed.) (1976). *Language Testing Symposium: A Psycholinguistic Approach*. London : Oxford University Press.
- Davies, A. and Widdowson, H.G. (1974). *Techniques in Applied Linguistics*. London : Oxford University Press.
- Devine, M. and Yaghlian, N. (2000). *Construction of an Objective Test*. Retrieved from <http://www.clt.cornell.edu/campus/teach/faculty/Materials/TestConstructionManual.pdf> on June 2, 2004.
- DIIA. (2003). *Test Item Analysis and Decision Making*. Division of Instructional Innovation and Assessment, University of Texas at Austin. <http://www.utexas.edu/academic/diia> on June 8, 2004.
- Downie, M. N. and Heath, R.W. (1970). *Basic Statistics*. New York: Harper and Row.
- Dunkel, P. A. (1999). "Considerations in developing or using second /foreign language proficiency computer-adaptive tests." *Language Learning & Technology* 2, 2, 77-93
- Ebel, R.L. (1972). *Essentials of Educational Measurement*. New Jersey: Englewood Cliffs, Prentice Hall.
- Genesee, F. and Upshur, J. (1996). *Classroom Evaluation in Second Language Education*. Cambridge University Press.
- Gipps, C. V. (1994). *Beyond testing: Towards a theory of educational assessment*. London: The Falmer Press.
- Gronlund, N. (1976). *Measurement and Evaluation in Teaching*, 3rd Ed. New York: Macmillan.
- Haladyna, T. M. (1992). "The effectiveness of several multiple-choice item formats." *Applied Measurement in Education*, 5, 73-88.
- Haladyna, T. M. (1994). *Developing and validating multiple-choice test items*. Hillsdale: Lawrence Erlbaum.
- Haladyna, T. M. and Downing, S. M. (1989). "A taxonomy of multiple-choice item writing rules." *Applied Measurement in Education*, 2, 37-50.
- Hambleton, R. K., Swaminathan, H. and Rogers, H. J. (1991). *Fundamentals of item response theory*. London: Sage Publications.
- Hambleton, R.K. (1980). "Test Score Validity and Standard Setting Methods." In A. Berk (ed.) *Criterion-referenced Measurement: The State of the Art*. Maryland: The Johns Hopkins University Press.
- Hambleton, R.K. and Cook, L.L. (1977). "Latent Trait Models and Their Use in the Analysis of Educational Test Data", *Journal of Educational Measurement*, 14(Summer).
- Heaton, J. B. (1988). *Writing English language tests*. London: Longman Group UK Limited.
- Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge: Cambridge University Press.

- Internet World Stat. (2006). *Internet Users by Language*. Retrieved from <http://www.internetworldstats.com/stats7.htm> on August 7, 2006.
- Jakobovits, L.A. (1970). *Foreign Language Learning: A Psycholinguistic Analysis of the Issues*. Mass : Newbury House.
- Jerard, K. (1995). *Basic Item Analysis for Multiple-Choice Tests*. ERIC Clearing House. Retrieved from http://www.eric.ed.gov/ERICDocs/data/ericdocs2/content_storage_01/0000000b/80/2a/26/4c.pdf on July 10, 2005.
- Kitao, S. K. and Kitao, K. (1996a). *Testing Communicative Competence*. *Internet TESOL Journal*, 2, 5.
- Kitao, S. K. and Kitao, K. (1996b). *Testing Grammar*. *Internet TESOL Journal*, 2, 6.
- Kitao, S. K. and Kitao, K. (1996c). *Testing Listening*. *Internet TESOL Journal*, 2, 7.
- Kunnan, A. J. (Ed.). (1998). *Validation in language assessment: Selected papers from the 17th Language Testing Research Colloquium*, Long Beach. London: LEA.
- Lado, R. (1977). *Language Testing*. London: Longman. (Ninth Edition)
- Linn, L. R. (ed.) (1989). *Educational measurement* (3rd ed.). London: Collier Macmillan Publishers.
- Lord, F.M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale: Lawrence Erlbaum.
- Madsen, H. S. (1983). *Techniques in testing*. Oxford: Oxford University Press.
- Marshall, J.C. and Hales, L.W. (1972). *Essentials of Testing*. Calif.: Addison-Wesley.
- McNamara, T. (1996). *Measuring second language performance*. London: Addison Wesley Longman Limited.
- McNamara, T. (2000). *Language testing*. Oxford: Oxford University Press.
- Miller, B.J. and Sundre, D.L. (2000). *Assessment in General Education: A Case Study in Scientific and Quantitative Reasoning*. Retrieved from <http://www.jmu.edu/assessment> on June 8, 2002.
- Morrow, K. (1977). *Techniques of Evaluation for a Notional Syllabus*. University of Reading.
- Osterlind, S. J (1989). *Constructing Test Items*. Londres: Kluwer Academic Publishers.
- Plake, B.S. and Others (1982). "Effects of Item Arrangement, Knowledge of Arrangement, Test Anxiety and Sex on Test Performance," *Journal of Educational Measurement*, Vol. 19.
- Rivers, W.M. (1972). "Testing and Student Learning" in M.C. O'Brien (ed.), *Testing in Second Language Teaching: New Dimensions*. Ireland, ATESOL.
- Roid, H. and Haladyna, T.(1982). *A Technology for Test Items Writing*. New York: Academic Press.
- Rudner, L.M. and Others. (1980). "A Monte Carlo Comparison of Seven Biased Item Detection Techniques." *Journal of Educational Measurement* 17 (Spring).
- Runte, N. (2001). *Item Analysis without complicated Statistics*. Retrieved from <http://www.uleth.ca/edu/runte/tests/itemanalysis.html> on July 8, 2005.
- Shavelson, R. J. and Webb, N. M. (1991). *Generalizability theory: A primer*. London: Sage Publications.
- Shepard, L.A.(1982). "Definitions of Bias" In . R.A. Berk (ed.) *Handbook of Methods for Detecting Test Bias*. Maryland: The Johns Hopkins University Press.
- Terwilliger, J. S. (1977). "Assigning Grades — Philosophical Issues and Practical Recommendation." *Journal of Educational Measurement* Vol. 15.
- Tinsley, H.A. and Davis, R. (1975). "An Investigation of the Rasch Simple Logistic Model: Sample Free Item and Test Calibration." *Journal of Educational and Psychological Measurement*, 35.

- University of Minnesota. (2002). *Visual Assessment Center*. Retrieved from <http://www.carla.umn.edu/assessment/VAC/research/construction.html> on June 7, 2002.
- Valette, R. M. (1969). *Directions in foreign language testing*. NY: Modern Language Association.
- Valtette, R.M. and Disick, R.S. (1972). *Modern Language Performance Objectives and Individualization*. New York: Harcourt Brace.
- Vistawide. (2006). *World Languages and Cultures*. Retrieved from http://www.vistawide.com/languages/language_statistics.htm on August 7, 2006.
- Warm, T.A. (1978). *Primer of Item Response Theory*. Oklahoma: US Coast Guard Institute.
- Weir, C. J. (1990). *Communicative language testing*. New York: Prentice Hall.
- Weir, C. J. and Others (eds.). (2000). *Studies in Language Testing*. Cambridge University Press.
- Weir, C. J. and Roberts, J. (1994). *Evaluation in ELT Pragmatics*. Oxford: Oxford University Press.
- Willmott, A.S. (1980). "What does Rasch Promise?," *Journal of NFER*, 22.
- Willmott, A.S. and Fowles, D.E. (1974). *The Objective interpretation of test performance: the Rasch model applied*. NFER Publishing Company Ltd.
- Wright, B.D. (1977). "Solving Measurement Problems with the Rasch Model." *Journal of Educational Measurement*, 14.
- Wright, B.D. and Mead (1977). *BICAL: Calibrating Items and Scales with the Rasch Model (Research Memorandum 23)*. Chicago, Illinois: Statistical Laboratory, Department of Education, University of Chicago.
- Wright, B.D. and Panchapakesan, N. (1969). "A Procedure for Sample-free Item Analysis." *Journal of Educational and Psychological Measurement*, 29.
- Wright, B.D. and Stone, M.H. (1979). *Best Test Design*. Chicago: Mesa Press.
- Zimmerman, B.B. and Others. (1990). *How to Prepare Better Tests*. Retrieved from http://www.idea.ksu.edu/papers/Idea_Paper_16.pdf on July 12, 2004.

ภาคผนวก ก. ตัวอย่างผลการวิเคราะห์ข้อทดสอบแบบประเพณีนิยม

CTIA 47

This is CLASSICAL TEST ITEM ANALYSIS PROGRAM, Version 8 (2547)

to analyze M/C item and assign grades.

It was originally written for DOS by

Dr. Suphat Sukamolson, Ph.D. *

(Educational Measurement and Evaluation),

and updated for Windows by

Mr. Komson Eiamjamrus, M.Sc.

LANGUAGE INSTITUTE

CHULALONGKORN UNIVERSITY, 2004

ALL RIGHTS RESERVED

[* Sukamolson, S. *Computerized Item Analyses and Grading*. BKK: Wittayaphat, 2542.]

CONTROL PARAMETERS**Total Items:** 100**Items Analyzed:** 100**From Item:** 1**To Item:** 100**Number of Choices:** 4**NUMERIC ANSWER KEYS****KEYS:** 11142322241313212333424241232222313421433141111334233433433444224321411312313443413431121
1443332243

11142	32224	13132	12333	42424	12322	22313	42143	31411	11334
23343	34333	44422	43214	11312	31344	34134	31121	14433	32243

RANKED SCORES

RANK SCORES	RAW SCORES	PERCENT SCORES	IDENTIFICATION
1	90	90.0	005
2	85	85.0	013
3	83	83.0	035
4	82	82.0	091
5	81	81.0	105
6	80	80.0	008
6	80	80.0	087
8	77	77.0	090
8	77	77.0	115
10	76	76.0	003
10	76	76.0	009
10	76	76.0	067
13	75	75.0	028
13	75	75.0	049
15	73	73.0	030
15	73	73.0	032
15	73	73.0	114
18	72	72.0	034
18	72	72.0	066
18	72	72.0	086
18	72	72.0	104
22	71	71.0	044
22	71	71.0	084
22	71	71.0	092
25	70	70.0	082
26	69	69.0	015
26	69	69.0	063
28	68	68.0	011
28	68	68.0	036
30	67	67.0	047
30	67	67.0	062
32	65	65.0	038
32	65	65.0	053
32	65	65.0	054
32	65	65.0	083
32	65	65.0	112
37	64	64.0	016

SCORE STATISTICS

No. of Items: 100
Respondents: 116
Mean Score: 57.87
Standard Deviation: 13.30
Mean Standard Error: 1.235
Maximum: 90
Minimum: 25
Range: 65
Quartile Deviation: 9.75
Median: 57.00
Mode*: 57
Skewness: -0.02
Kurtosis: -0.44

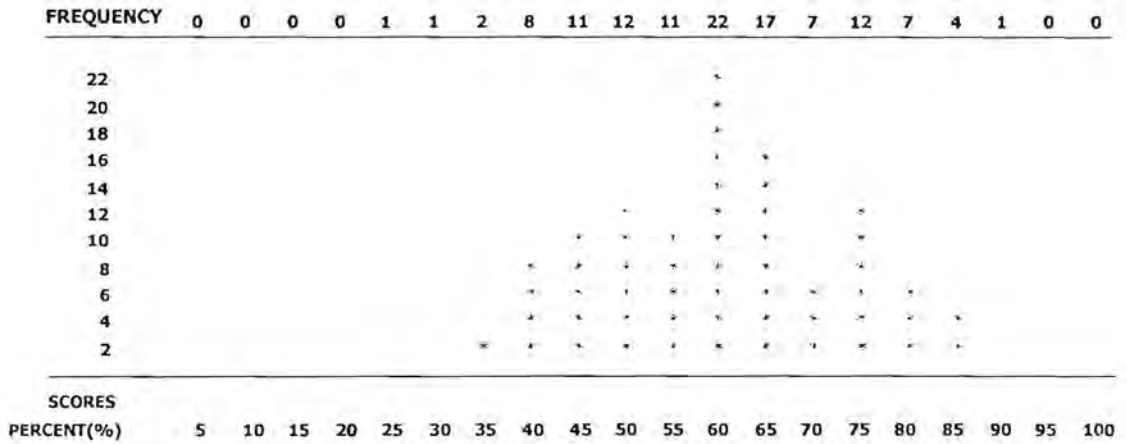
*** Estimated Mode,**

**If the score distribution is not normal,
look for the actual mode.
(The score with highest frequency.)**

DISTRIBUTION OF SCORES

SCORES		CUMULATIVE								
RAW	PERCENT	FREQUENCY	CUMULATIVE FREQUENCY	PERCENT FREQUENCY	PERCENT FREQUENCY	PERCENTILE RANK	NORMALIZED T-SCORE	STANINE SCORE	Z SCORE	LINEAR T-SCORE
90	90.0	1	1	0.9	0.9	99.6	77	1	2.415	74.2
85	85.0	1	2	0.9	1.7	98.7	73	1	2.040	70.4
83	83.0	1	3	0.9	2.6	97.8	71	1	1.889	68.9
82	82.0	1	4	0.9	3.4	97.0	69	1	1.814	68.1
81	81.0	1	5	0.9	4.3	96.1	68	1	1.739	67.4
80	80.0	2	7	1.7	6.0	94.8	67	2	1.664	66.6
77	77.0	2	9	1.7	7.8	93.1	65	2	1.438	64.4
76	76.0	3	12	2.6	10.3	90.9	64	2	1.363	63.6
75	75.0	2	14	1.7	12.1	88.8	63	3	1.288	62.9
73	73.0	3	17	2.6	14.7	86.6	62	3	1.137	61.4
72	72.0	4	21	3.4	18.1	83.6	60	3	1.062	60.6
71	71.0	3	24	2.6	20.7	80.6	59	3	0.987	59.9
70	70.0	1	25	0.9	21.6	78.9	59	3	0.912	59.1
69	69.0	2	27	1.7	23.3	77.6	58	3	0.837	58.4
68	68.0	2	29	1.7	25.0	75.9	58	4	0.762	57.6
67	67.0	2	31	1.7	26.7	74.1	57	4	0.686	56.9
65	65.0	5	36	4.3	31.0	71.1	56	4	0.536	55.4
64	64.0	3	39	2.6	33.6	67.7	55	4	0.461	54.6
63	63.0	2	41	1.7	35.3	65.5	54	4	0.386	53.9
62	62.0	3	44	2.6	37.9	63.4	54	4	0.310	53.1
61	61.0	4	48	3.4	41.4	60.3	53	4	0.235	52.4
60	60.0	3	51	2.6	44.0	57.3	52	5	0.160	51.6
59	59.0	4	55	3.4	47.4	54.3	52	5	0.085	50.8
58	58.0	2	57	1.7	49.1	51.7	51	5	0.010	50.1
57	57.0	8	65	6.9	56.0	47.4	50	5	-0.065	49.3
56	56.0	5	70	4.3	60.3	41.8	48	5	-0.141	48.6
55	55.0	5	75	4.3	64.7	37.5	47	6	-0.216	47.8
53	53.0	4	79	3.4	68.1	33.6	46	6	-0.366	46.3
52	52.0	1	80	0.9	69.0	31.5	46	6	-0.441	45.6
51	51.0	1	81	0.9	69.8	30.6	45	6	-0.517	44.8
50	50.0	2	83	1.7	71.6	29.3	45	6	-0.592	44.1
48	48.0	6	89	5.2	76.7	25.9	44	6	-0.742	42.6
47	47.0	2	91	1.7	78.4	22.4	43	7	-0.817	41.8
46	46.0	2	93	1.7	80.2	20.7	42	7	-0.892	41.1
45	45.0	2	95	1.7	81.9	19.0	42	7	-0.968	40.3
44	44.0	1	96	0.9	82.8	17.7	41	7	-1.043	39.6
43	43.0	3	99	2.6	85.3	15.9	41	7	-1.118	38.8
42	42.0	3	102	2.6	87.9	13.4	39	7	-1.193	38.1
41	41.0	2	104	1.7	89.7	11.2	38	7	-1.268	37.3
40	40.0	2	106	1.7	91.4	9.5	37	8	-1.344	36.6
39	39.0	2	108	1.7	93.1	7.8	36	8	-1.419	35.8
38	38.0	3	111	2.6	95.7	5.6	35	8	-1.494	35.1
36	36.0	1	112	0.9	96.6	3.9	33	9	-1.644	33.6
33	33.0	1	113	0.9	97.4	3.0	32	9	-1.870	31.3

HISTOGRAM OF SCORE DISTRIBUTION



ITEM ANALYSIS

ITEM NO.	RESP-ONSE	NUMBER RESPONDING			DIFFICULTY INDICE				CORRELATION COEFFICIENTS					MEAN CRITERION	
		UPPER	MIDDLE	LOWER	UPPER	LOWER	TOTAL	DELTA	DISCINDEX	BISERIAL RBIS	POINT-BISERIAL t	RPB	t	SCORE	T-SC
1	* 1	15	23	13	0.484	0.419	0.440	13.65	0.065	0.079	0.85	0.063	0.68	58.8	50.72
	2	0	4	9	0.000	0.290	0.112	17.90	-0.290	-0.482	-5.88	-0.319	-3.60	45.9	41.02
	3	16	26	8	0.516	0.258	0.431	13.73	0.258	0.190	2.07	0.152	1.65	60.2	51.75
	4	0	1	1	0.000	0.032	0.017	21.50	-0.032	-0.124	-1.34	-0.049	-0.52	53.0	46.34
	ERROR	0	0	0	0.000	0.000	0.000	0.00	0.000	0.000	0.00	0.000	0.00	0.0	0.00
2	* 1	25	18	8	0.806	0.258	0.440	13.65	0.548	0.500	6.17	0.400	4.66	63.9	54.52
	2	0	6	2	0.000	0.065	0.069	18.97	-0.065	-0.097	-1.05	-0.059	-0.63	55.0	47.84
	3	4	15	11	0.129	0.355	0.259	15.63	-0.226	-0.239	-2.62	-0.182	-1.98	53.8	46.91
	4	1	15	10	0.032	0.323	0.224	16.07	-0.290	-0.371	-4.26	-0.279	-3.10	51.0	44.81
	ERROR	1	0	0	0.032	0.000	0.009	22.56	0.032	0.416	4.89	0.134	1.44	77.0	64.38
3	* 1	10	9	5	0.323	0.161	0.207	16.31	0.161	0.216	2.37	0.162	1.75	62.1	53.17
	2	0	3	3	0.000	0.097	0.052	19.55	-0.097	-0.386	-4.47	-0.214	-2.34	45.7	40.83
	3	19	35	20	0.613	0.645	0.638	11.55	-0.032	-0.008	-0.08	-0.006	-0.06	57.8	49.95
	4	2	7	3	0.065	0.097	0.103	18.09	-0.032	-0.077	-0.83	-0.050	-0.53	55.9	48.53
	ERROR	0	0	0	0.000	0.000	0.000	0.00	0.000	0.000	0.00	0.000	0.00	0.0	0.00
4	1	0	4	1	0.000	0.032	0.043	19.90	-0.032	-0.088	-0.95	-0.046	-0.49	55.0	47.84
	2	0	2	4	0.000	0.129	0.052	19.55	-0.129	-0.370	-4.25	-0.206	-2.24	46.2	41.20
	3	1	4	3	0.032	0.097	0.069	18.97	-0.065	-0.182	-1.98	-0.110	-1.18	52.5	45.96
	* 4	30	44	23	0.968	0.742	0.836	9.05	0.226	0.313	3.52	0.223	2.45	59.2	50.99
	ERROR	0	0	0	0.000	0.000	0.000	0.00	0.000	0.000	0.00	0.000	0.00	0.0	0.00
5	1	0	0	7	0.000	0.226	0.060	19.25	-0.226	-0.793	-13.90	-0.441	-5.25	34.7	32.59
	* 2	30	53	18	0.968	0.581	0.871	8.44	0.387	0.713	10.87	0.483	5.89	60.3	51.86
	3	0	1	5	0.000	0.161	0.052	19.55	-0.161	-0.449	-5.37	-0.249	-2.75	43.7	39.32
	4	0	0	1	0.000	0.032	0.009	22.56	-0.032	-0.280	-3.12	-0.090	-0.97	45.0	40.32
	ERROR	1	0	0	0.032	0.000	0.009	22.56	0.032	0.220	2.41	0.071	0.76	68.0	57.62
6	1	1	3	7	0.032	0.226	0.095	18.29	-0.194	-0.278	-3.09	-0.176	-1.91	50.6	44.56
	2	3	6	8	0.097	0.258	0.147	17.41	-0.161	-0.298	-3.33	-0.196	-2.13	51.6	45.28
	* 3	20	28	8	0.645	0.258	0.483	13.21	0.387	0.319	3.59	0.255	2.81	61.4	52.63
	4	7	17	8	0.226	0.258	0.276	15.42	-0.032	-0.019	-0.20	-0.014	-0.15	57.6	49.77
	ERROR	0	0	0	0.000	0.000	0.000	0.00	0.000	0.000	0.00	0.000	0.00	0.0	0.00
7	1	4	8	9	0.129	0.290	0.181	16.68	-0.161	-0.257	-2.84	-0.187	-2.04	52.6	46.02
	* 2	27	45	21	0.871	0.677	0.802	9.57	0.194	0.327	3.69	0.236	2.59	59.4	51.17
	3	0	0	0	0.000	0.000	0.000	0.00	0.000	0.000	0.00	0.000	0.00	0.0	0.00
	4	0	1	1	0.000	0.032	0.017	21.50	-0.032	-0.430	-5.09	-0.168	-1.82	41.0	37.32
	ERROR	0	0	0	0.000	0.000	0.000	0.00	0.000	0.000	0.00	0.000	0.00	0.0	0.00
8	1	2	17	11	0.065	0.355	0.259	15.63	-0.290	-0.289	-3.22	-0.221	-2.42	52.9	46.26
	* 2	18	10	8	0.581	0.258	0.310	15.02	0.323	0.250	2.76	0.196	2.13	61.8	52.92
	3	1	4	4	0.032	0.129	0.078	18.73	-0.097	-0.202	-2.20	-0.123	-1.33	52.2	45.75
	4	10	23	8	0.323	0.258	0.353	14.54	0.065	0.102	1.10	0.082	0.88	59.3	51.11
	ERROR	0	0	0	0.000	0.000	0.000	0.00	0.000	0.000	0.00	0.000	0.00	0.0	0.00

TEST SUMMARY**TEST STATISTICS**

	Mean	Min	Median*	Max	Std Devn	Var
Test Scores	57.871	25.000	57.500	90.000	13.302	176.931
Diff. Index	0.579	0.164	0.556	0.948	0.503	0.253
Delta	12.029	6.448	11.701	16.954	2.349	5.519
Disc. Index	0.334	0.000	0.274	0.548	0.167	0.028
Biserial (RBIS)	0.383	0.002	0.369	0.736	0.162	0.026
Point-Biserial (RPB)	0.292	0.002	0.266	0.531	0.124	0.015

KUDER-RICHARDSON RELIABILITY STATISTICS

KR20 = 0.893 SEM20 = 4.357

KR21 = 0.871 SEM21 = 4.779

CRONBACH ALPHA RELIABILITY STATISTICS

ALPHA = 0.893 SEM-ALP = 4.357

SPLIT-HALF RELIABILITY STATISTICS

RTT = 0.865 SEMTT = 4.883

* Approximate medians,
if the distributions are not normal.

DISTRIBUTION OF DIFFICULTIES

Plotted over 20 equal intervals of 0.039

Min	Median*																		Max	
0.163	0.203	0.242	0.281	0.321	0.360	0.399	0.438	0.478	0.517	0.556	0.595	0.634	0.674	0.713	0.752	0.791	0.831	0.870	0.909	0.949
76	3	0	8	47	11	17	1	6	12	37	22	9	24	30	89	7	4	5	50	
0	72	0	21	68	60	20	2	19	13	41	73	36	27	32	0	23	14	10	81	
0	74	0	38	0	66	26	28	25	15	43	97	39	33	35	0	31	18	29	82	
0	78	0	42	0	80	49	46	44	16	52	0	53	40	45	0	34	95	0	88	
0	0	0	54	0	0	70	48	51	75	58	0	61	56	77	0	62	0	0	90	
0	0	0	55	0	0	87	57	71	84	69	0	63	59	0	0	100	0	0	92	
0	0	0	64	0	0	0	79	91	94	83	0	65	85	0	0	0	0	0	0	
0	0	0	67	0	0	0	96	0	0	99	0	86	93	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	98	0	0	0	0	0	0	0	
Col 1	4	0	8	2	4	6	8	7	7	8	3	9	8	5	1	6	4	3	6	
Totals																				

**** If any index is zero,
its item no. is not plotted.**

DISTRIBUTION OF DISCRIMINATIONS

Plotted over 20 equal intervals of 0.035

Min	Median*																		Max	
	-0.001	0.071	0.142	0.213	0.284	0.355	0.426	0.497	0.568	0.639	0.711									
	0.035	0.106	0.177	0.248	0.319	0.390	0.461	0.532	0.603	0.674										
56	1	47	23	3	7	4	13	10	8	5	18	44	11	24	2	58	36	35	75	
0	66	92	90	28	25	53	26	14	9	6	20	46	16	34	17	73	98	71	97	
0	96	0	0	37	50	54	29	19	12	31	27	0	32	39	43	0	0	0	99	
0	0	0	0	74	62	69	30	49	15	38	57	0	51	55	0	0	0	0	0	
0	0	0	0	82	64	76	40	60	22	41	65	0	77	59	0	0	0	0	0	
0	0	0	0	88	70	0	45	61	33	42	95	0	83	63	0	0	0	0	0	
0	0	0	0	0	84	0	72	67	52	48	0	0	94	86	0	0	0	0	0	
0	0	0	0	0	87	0	80	89	68	85	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	91	0	79	93	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	
Col	1	3	2	2	6	8	5	9	8	10	9	6	2	7	7	3	2	2	2	3
Totals																				

** If any index is zero,
its item no. is not plotted.

CTIA 47

This is CLASSICAL TEST ITEM ANALYSIS PROGRAM, Version 8 (2547)

to analyze M/C item and assign grades.

It was originally written for DOS by

Dr. Suphat Sukamolson, Ph.D. *

(Educational Measurement and Evaluation),

and updated for Windows by

Mr. Komson Eiamjamrus, M.Sc.

LANGUAGE INSTITUTE

CHULALONGKORN UNIVERSITY, 2004

ALL RIGHTS RESERVED

[* Sukamolson, S. *Computerized Item Analyses and Grading*. BKK: Wittayaphat, 2542.]

CONTROL PARAMETERS**Total Items:** 100**Items Analyzed:** 100**From Item:** 1**To Item:** 100**Number of Choices:** 4**NUMERIC ANSWER KEYS****KEYS:** 11134224333233112112231332322313312232413341111312421432434212421331112121441233312314422
32322244111134 22433 33233 11211 22313 32322 31331 22324 13341 11131
24214 32434 21242 13311 12121 44123 33123 14422 32322 22441

RANKED SCORES

RANK SCORES	RAW SCORES	PERCENT SCORES	IDENTIFICATION
1	92	92.0	043
2	90	90.0	101
3	86	86.0	024
3	86	86.0	031
5	83	83.0	063
5	83	83.0	102
7	79	79.0	057
8	78	78.0	015
8	78	78.0	040
8	78	78.0	058
11	77	77.0	005
11	77	77.0	012
11	77	77.0	032
11	77	77.0	098
15	76	76.0	093
16	75	75.0	016
16	75	75.0	020
18	74	74.0	028
19	73	73.0	025
19	73	73.0	110
19	73	73.0	121
22	72	72.0	010
22	72	72.0	042
22	72	72.0	124
25	71	71.0	023
25	71	71.0	064
25	71	71.0	095
28	69	69.0	038
28	69	69.0	099
28	69	69.0	117
31	68	68.0	026
31	68	68.0	066
31	68	68.0	103
34	67	67.0	011
34	67	67.0	053
36	66	66.0	004
36	66	66.0	018

SCORE STATISTICS

No. of Items: 100
Respondents: 124
Mean Score: 55.32
Standard Deviation: 16.15
Mean Standard Error: 1.451
Maximum: 92
Minimum: 25
Range: 67
Quartile Deviation: 12.50
Median: 55.00
Mode*: 47
Skewness: 0.09
Kurtosis: -0.88

* Estimated Mode,

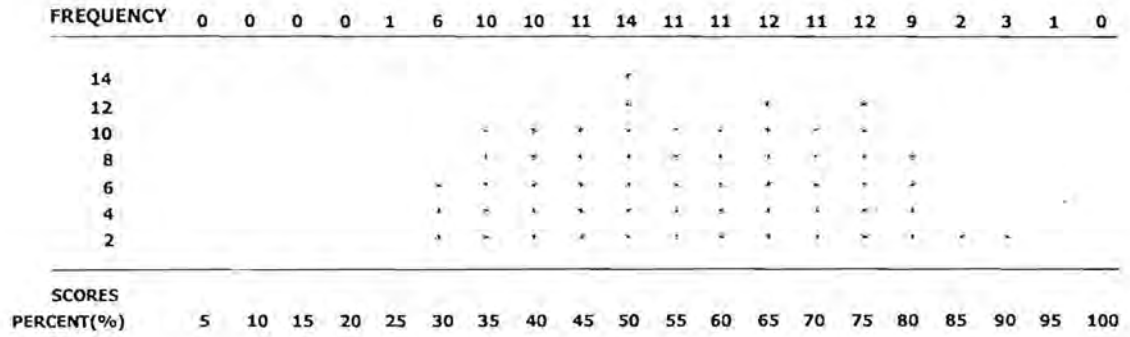
If the score distribution is not normal,
look for the actual mode.

(The score with highest frequency.)

DISTRIBUTION OF SCORES

SCORES		FREQUENCY	CUMULATIVE				NORMALIZED T-SCORE	STANINE SCORE	Z SCORE	LINEAR T-SCORE
RAW	PERCENT		FREQUENCY	PERCENT FREQUENCY	PERCENT FREQUENCY	PERCENTILE RANK				
92	92.0	1	1	0.8	0.8	99.6	77	1	2.271	72.7
90	90.0	1	2	0.8	1.6	98.8	73	1	2.147	71.5
86	86.0	2	4	1.6	3.2	97.6	70	1	1.899	69.0
83	83.0	2	6	1.6	4.8	96.0	68	2	1.713	67.1
79	79.0	1	7	0.8	5.6	94.8	67	2	1.466	64.7
78	78.0	3	10	2.4	8.1	93.1	65	2	1.404	64.0
77	77.0	4	14	3.2	11.3	90.3	64	2	1.342	63.4
76	76.0	1	15	0.8	12.1	88.3	62	3	1.280	62.8
75	75.0	2	17	1.6	13.7	87.1	62	3	1.218	62.2
74	74.0	1	18	0.8	14.5	85.9	61	3	1.156	61.6
73	73.0	3	21	2.4	16.9	84.3	61	3	1.094	60.9
72	72.0	3	24	2.4	19.4	81.9	60	3	1.032	60.3
71	71.0	3	27	2.4	21.8	79.4	59	3	0.971	59.7
69	69.0	3	30	2.4	24.2	77.0	58	3	0.847	58.5
68	68.0	3	33	2.4	26.6	74.6	57	4	0.785	57.8
67	67.0	2	35	1.6	28.2	72.6	57	4	0.723	57.2
66	66.0	3	38	2.4	30.6	70.6	56	4	0.661	56.6
65	65.0	2	40	1.6	32.3	68.5	55	4	0.599	56.0
64	64.0	3	43	2.4	34.7	66.5	55	4	0.537	55.4
63	63.0	1	44	0.8	35.5	64.9	54	4	0.475	54.8
62	62.0	2	46	1.6	37.1	63.7	54	4	0.413	54.1
61	61.0	4	50	3.2	40.3	61.3	53	4	0.351	53.5
59	59.0	5	55	4.0	44.4	57.7	52	5	0.228	52.3
58	58.0	2	57	1.6	46.0	54.8	52	5	0.166	51.7
57	57.0	2	59	1.6	47.6	53.2	51	5	0.104	51.0
56	56.0	2	61	1.6	49.2	51.6	51	5	0.042	50.4
55	55.0	2	63	1.6	50.8	50.0	50	5	-0.020	49.8
54	54.0	3	66	2.4	53.2	48.0	50	5	-0.082	49.2
52	52.0	4	70	3.2	56.5	45.2	49	5	-0.206	47.9
51	51.0	2	72	1.6	58.1	42.7	49	5	-0.268	47.3
50	50.0	2	74	1.6	59.7	41.1	48	5	-0.330	46.7
49	49.0	4	78	3.2	62.9	38.7	48	6	-0.391	46.1
48	48.0	1	79	0.8	63.7	36.7	47	6	-0.453	45.5
47	47.0	6	85	4.8	68.5	33.9	46	6	-0.515	44.8
46	46.0	1	86	0.8	69.4	31.0	46	6	-0.577	44.2
45	45.0	2	88	1.6	71.0	29.8	45	6	-0.639	43.6
44	44.0	3	91	2.4	73.4	27.8	45	6	-0.701	43.0
43	43.0	3	94	2.4	75.8	25.4	44	6	-0.763	42.4
42	42.0	3	97	2.4	78.2	23.0	43	7	-0.825	41.8
40	40.0	1	98	0.8	79.0	21.4	43	7	-0.949	40.5
39	39.0	3	101	2.4	81.5	19.8	42	7	-1.010	39.9
38	38.0	2	103	1.5	83.1	17.7	41	7	-1.072	39.3
37	37.0	2	105	1.6	84.7	16.1	41	7	-1.134	38.7
36	36.0	2	107	1.6	86.3	14.5	40	7	-1.196	38.0

HISTOGRAM OF SCORE DISTRIBUTION



ITEM ANALYSIS

ITEM NO.	RESP-ONSE	NUMBER RESPONDING			DIFFICULTY INDICE				CORRELATION COEFFICIENTS					MEAN CRITERION	
		UPPER	MIDDLE	LOWER	UPPER	LOWER	TOTAL	DELTA	DISCINDEX	BISERIAL RBIS	POINT-BISERIAL t	RPB	t	SCORE	T-SC
1	* 1	29	42	10	0.879	0.303	0.653	11.38	0.576	0.578	7.83	0.455	5.65	60.7	53.32
	2	0	2	7	0.000	0.212	0.073	18.87	-0.212	-0.477	-5.99	-0.281	-3.23	39.1	39.96
	3	2	8	10	0.061	0.303	0.161	16.99	-0.242	-0.345	-4.06	-0.252	-2.87	46.1	44.26
	4	2	6	6	0.061	0.182	0.113	17.88	-0.121	-0.245	-2.79	-0.162	-1.81	48.0	45.47
	ERROR	0	0	0	0.000	0.000	0.000	0.00	0.000	0.000	0.00	0.000	0.00	0.0	0.00
2	* 1	11	13	3	0.333	0.091	0.218	16.16	0.242	0.304	3.53	0.231	2.63	62.4	54.39
	2	2	25	16	0.061	0.485	0.347	14.62	-0.424	-0.421	-5.13	-0.332	-3.89	48.0	45.44
	3	17	10	4	0.515	0.121	0.250	15.74	0.394	0.452	5.60	0.349	4.12	65.1	56.05
	4	3	10	10	0.091	0.303	0.185	16.62	-0.212	-0.316	-3.68	-0.228	-2.59	47.6	45.22
	ERROR	0	0	0	0.000	0.000	0.000	0.00	0.000	0.000	0.00	0.000	0.00	0.0	0.00
3	* 1	29	40	9	0.879	0.273	0.629	11.64	0.606	0.570	7.66	0.456	5.65	61.0	53.50
	2	1	9	12	0.030	0.364	0.177	16.74	-0.333	-0.410	-4.96	-0.301	-3.48	44.9	43.53
	3	1	6	8	0.030	0.242	0.121	17.72	-0.212	-0.420	-5.12	-0.285	-3.28	42.9	42.33
	4	2	3	4	0.061	0.121	0.073	18.87	-0.061	-0.081	-0.90	-0.048	-0.53	52.6	48.29
	ERROR	0	0	0	0.000	0.000	0.000	0.00	0.000	0.000	0.00	0.000	0.00	0.0	0.00
4	1	2	5	3	0.061	0.091	0.081	18.64	-0.030	-0.129	-1.44	-0.077	-0.86	51.1	47.39
	2	2	12	7	0.061	0.212	0.169	16.87	-0.152	-0.178	-2.00	-0.127	-1.42	50.8	47.18
	* 3	24	19	5	0.727	0.152	0.387	14.19	0.576	0.538	7.05	0.432	5.29	64.1	55.44
	4	5	22	18	0.152	0.545	0.363	14.44	-0.394	-0.371	-4.41	-0.294	-3.40	49.0	46.10
	ERROR	0	0	0	0.000	0.000	0.000	0.00	0.000	0.000	0.00	0.000	0.00	0.0	0.00
5	1	1	3	1	0.030	0.030	0.040	20.03	0.000	0.048	0.54	0.024	0.26	57.2	51.16
	2	0	5	2	0.000	0.061	0.056	19.38	-0.061	-0.182	-2.04	-0.104	-1.16	48.4	45.73
	3	13	23	22	0.394	0.667	0.468	13.36	-0.273	-0.245	-2.79	-0.196	-2.21	51.9	47.91
	* 4	19	27	8	0.576	0.242	0.435	13.69	0.333	0.295	3.41	0.236	2.68	59.7	52.69
	ERROR	0	0	0	0.000	0.000	0.000	0.00	0.000	0.000	0.00	0.000	0.00	0.0	0.00
6	1	0	4	2	0.000	0.061	0.048	19.68	-0.061	-0.210	-2.38	-0.112	-1.24	47.3	45.05
	* 2	24	24	16	0.727	0.485	0.516	12.80	0.242	0.248	2.83	0.198	2.23	58.4	51.92
	3	2	5	2	0.061	0.061	0.073	18.87	0.000	-0.085	-0.94	-0.050	-0.55	52.4	48.22
	4	7	25	13	0.212	0.394	0.363	14.44	-0.182	-0.163	-1.82	-0.129	-1.44	52.6	48.29
	ERROR	0	0	0	0.000	0.000	0.000	0.00	0.000	0.000	0.00	0.000	0.00	0.0	0.00
7	1	2	3	1	0.061	0.030	0.048	19.68	0.030	-0.008	-0.09	-0.005	-0.05	55.0	49.80
	* 2	29	40	19	0.879	0.576	0.710	10.75	0.303	0.383	4.59	0.297	3.43	58.4	51.90
	3	2	12	9	0.061	0.273	0.185	16.62	-0.212	-0.289	-3.34	-0.209	-2.36	48.3	45.63
	4	0	3	4	0.000	0.121	0.056	19.38	-0.121	-0.397	-4.77	-0.228	-2.58	40.3	40.69
	ERROR	0	0	0	0.000	0.000	0.000	0.00	0.000	0.000	0.00	0.000	0.00	0.0	0.00
8	1	2	7	15	0.061	0.455	0.194	16.50	-0.394	-0.462	-5.75	-0.341	-4.00	44.1	43.04
	2	0	3	5	0.000	0.152	0.065	19.11	-0.152	-0.388	-4.64	-0.225	-2.55	41.5	41.44
	3	1	3	5	0.030	0.152	0.073	18.87	-0.121	-0.317	-3.69	-0.186	-2.10	44.6	43.33
	* 4	30	45	8	0.909	0.242	0.669	11.21	0.667	0.636	9.10	0.506	6.49	61.1	53.56
	ERROR	0	0	0	0.000	0.000	0.000	0.00	0.000	0.000	0.00	0.000	0.00	0.0	0.00

TEST SUMMARY**TEST STATISTICS**

	Mean	Min	Median*	Max	Std Devn	Var
Test Scores	55.323	25.000	58.500	92.000	16.154	260.936
Diff. Index	0.553	0.105	0.512	0.919	0.507	0.257
Delta	12.376	7.358	12.707	18.057	2.354	5.543
Disc. Index	0.405	-0.303	0.273	0.848	0.211	0.045
Biserial (RBIS)	0.459	-0.328	0.256	0.841	0.201	0.040
Point-Biserial (RPB)	0.353	-0.258	0.207	0.672	0.157	0.025

KUDER-RICHARDSON RELIABILITY STATISTICS

KR20 = 0.930 SEM20 = 4.261

KR21 = 0.914 SEM21 = 4.726

CRONBACH ALPHA RELIABILITY STATISTICS

ALPHA = 0.930 SEM-ALP = 4.261

SPLIT-HALF RELIABILITY STATISTICS

RTT = 0.925 SEMTT = 4.416

* Approximate medians,
if the distributions are not normal.

DISTRIBUTION OF DIFFICULTIES

Plotted over 20 equal intervals of 0.041

Min	Median*																		Max
	0.104	0.186	0.268	0.349	0.431	0.512	0.594	0.675	0.756	0.838	0.920								
	0.146	0.227	0.308	0.390	0.471	0.553	0.634	0.716	0.797	0.879									
50	58	2	14	32	28	4	11	5	15	6	43	3	1	7	36	33	26	38	82
0	0	46	22	45	77	12	20	17	16	29	48	30	8	13	40	37	27	39	85
0	0	69	61	79	0	31	23	18	19	34	59	35	9	21	49	41	87	84	86
0	0	72	64	0	0	68	65	51	0	44	63	56	10	25	60	55	88	98	0
0	0	73	78	0	0	0	75	67	0	54	66	0	24	42	83	81	0	100	0
0	0	0	80	0	0	0	0	74	0	62	71	0	53	47	93	89	0	0	0
0	0	0	94	0	0	0	0	95	0	70	90	0	0	52	0	92	0	0	0
0	0	0	0	0	0	0	0	0	0	76	91	0	0	57	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	99	97	0	0	96	0	0	0	0	0
Col 1	1	5	7	3	2	4	5	7	3	9	9	4	6	9	6	7	4	5	3
Totals																			

**** If any index is zero,
its item no. is not plotted.**

160

DISTRIBUTION OF DISCRIMINATIONS

Plotted over 20 equal intervals of 0.058

Min	Median*																		Max
-0.304	-0.188	-0.073	0.042	0.158	0.273	0.388	0.503	0.618	0.733	0.849									
-0.245	-0.130	-0.015	0.100	0.215	0.330	0.445	0.561	0.676	0.791										
45	0	0	11	46	0	0	55	15	2	7	5	20	13	12	1	8	29	95	62
0	0	0	80	0	0	0	67	18	6	17	26	37	19	35	3	9	63	0	0
0	0	0	0	0	0	0	69	22	39	25	60	43	31	47	4	16	65	0	0
0	0	0	0	0	0	0	0	49	40	27	74	51	38	52	10	21	76	0	0
0	0	0	0	0	0	0	0	50	72	32	84	56	42	54	23	24	97	0	0
0	0	0	0	0	0	0	0	58	77	36	99	73	44	57	30	28	0	0	0
0	0	0	0	0	0	0	0	79	81	48	0	87	53	70	33	34	0	0	0
0	0	0	0	0	0	0	0	85	82	61	0	0	64	75	41	66	0	0	0
0	0	0	0	0	0	0	0	90	86	68	0	0	92	78	59	71	0	0	0
0	0	0	0	0	0	0	0	0	94	91	0	0	0	88	83	0	0	0	0
0	0	0	0	0	0	0	0	0	96	0	0	0	0	0	89	0	0	0	0
0	0	0	0	0	0	0	0	0	98	0	0	0	0	0	93	0	0	0	0
0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0
Col 1	0	0	2	1	0	0	3	9	13	10	6	7	9	10	12	9	5	1	1
Totals																			

**** If any index is zero,
its item no. is not plotted.**

CTIA 47

This is **CLASSICAL TEST ITEM ANALYSIS PROGRAM, Version 8 (2547)**

to analyze M/C item and assign grades.

It was originally written for DOS by

Dr. Suphat Sukamolson, Ph.D. *

(Educational Measurement and Evaluation),

and updated for Windows by

Mr. Komson Eiamjamrus, M.Sc.

LANGUAGE INSTITUTE

CHULALONGKORN UNIVERSITY, 2004

ALL RIGHTS RESERVED

[* Sukamolson, S. *Computerized Item Analyses and Grading*. BKK: Wittayaphat, 2542.]

CONTROL PARAMETERS**Total Items:** 100**Items Analyzed:** 100**From Item:** 1**To Item:** 100**Number of Choices:** 4**NUMERIC ANSWER KEYS****KEYS:** 344133324422121232311423123313341332442423111143132214114441213343131214141241221324141133
3342432411

34413	33244	22121	23231	14231	23313	34133	24424	23111	14313
22141	14441	21334	31312	14141	24122	13241	41133	33424	32411

RANKED SCORES

RANK SCORES	RAW SCORES	PERCENT SCORES	IDENTIFICATION
1	93	93.0	067
2	88	88.0	082
3	87	87.0	076
4	85	85.0	107
5	84	84.0	093
6	82	82.0	038
7	81	81.0	105
8	80	80.0	085
9	78	78.0	027
9	78	78.0	109
11	77	77.0	075
12	76	76.0	072
12	76	76.0	094
14	75	75.0	110
15	74	74.0	119
16	72	72.0	108
17	71	71.0	028
17	71	71.0	042
17	71	71.0	081
17	71	71.0	083
21	70	70.0	073
22	69	69.0	114
23	68	68.0	012
23	68	68.0	052
23	68	68.0	068
23	68	68.0	078
27	67	67.0	044
27	67	67.0	115
29	66	66.0	002
29	66	66.0	009
29	66	66.0	079
29	66	66.0	080
33	65	65.0	084
33	65	65.0	102
35	64	64.0	016
36	63	63.0	039
36	63	63.0	071

SCORE STATISTICS**No. of Items:** 100**Respondents:** 120**Mean Score:** 53.69**Standard Deviation:** 16.20**Mean Standard Error:** 1.479**Maximum:** 93**Minimum:** 22**Range:** 71**Quartile Deviation:** 12.25**Median:** 53.00**Mode*:** 47 or 43**Skewness:** 0.19**Kurtosis:** -0.69*** Estimated Mode,**

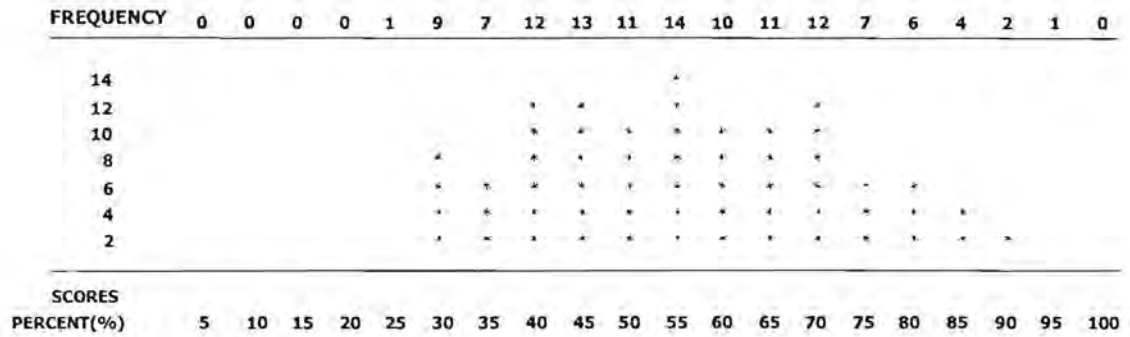
**If the score distribution is not normal,
look for the actual mode.**

(The score with highest frequency.)

DISTRIBUTION OF SCORES

SCORES		CUMULATIVE								
RAW	PERCENT	FREQUENCY	CUMULATIVE FREQUENCY	PERCENT FREQUENCY	PERCENT FREQUENCY	PERCENTILE RANK	NORMALIZED T-SCORE	STANINE SCORE	Z SCORE	LINEAR T-SCORE
93	93.0	1	1	0.8	0.8	99.6	77	1	2.426	74.3
88	88.0	1	2	0.8	1.7	98.8	73	1	2.118	71.2
87	87.0	1	3	0.8	2.5	97.9	71	1	2.056	70.6
85	85.0	1	4	0.8	3.3	97.1	69	1	1.932	69.3
84	84.0	1	5	0.8	4.2	96.3	68	1	1.871	68.7
82	82.0	1	6	0.8	5.0	95.4	67	2	1.747	67.5
81	81.0	1	7	0.8	5.8	94.6	67	2	1.686	66.9
80	80.0	1	8	0.8	6.7	93.8	66	2	1.624	66.2
78	78.0	2	10	1.7	8.3	92.5	65	2	1.500	65.0
77	77.0	1	11	0.8	9.2	91.3	64	2	1.439	64.4
76	76.0	2	13	1.7	10.8	90.0	63	2	1.377	63.8
75	75.0	1	14	0.8	11.7	88.8	63	3	1.315	63.2
74	74.0	1	15	0.8	12.5	87.9	62	3	1.253	62.5
72	72.0	1	16	0.8	13.3	87.1	62	3	1.130	61.3
71	71.0	4	20	3.3	16.7	85.0	61	3	1.068	60.7
70	70.0	1	21	0.8	17.5	82.9	60	3	1.007	60.1
69	69.0	1	22	0.8	18.3	82.1	60	3	0.945	59.4
68	68.0	4	26	3.3	21.7	80.0	59	3	0.883	58.8
67	67.0	2	28	1.7	23.3	77.5	58	3	0.821	58.2
66	66.0	4	32	3.3	26.7	75.0	57	4	0.760	57.6
65	65.0	2	34	1.7	28.3	72.5	56	4	0.698	57.0
64	64.0	1	35	0.8	29.2	71.3	56	4	0.636	56.4
63	63.0	3	38	2.5	31.7	69.6	56	4	0.575	55.7
62	62.0	3	41	2.5	34.2	67.1	55	4	0.513	55.1
61	61.0	2	43	1.7	35.8	65.0	54	4	0.451	54.5
60	60.0	3	46	2.5	38.3	62.9	54	4	0.389	53.9
59	59.0	2	48	1.7	40.0	60.8	53	4	0.328	53.3
58	58.0	1	49	0.8	40.8	59.6	53	5	0.266	52.7
57	57.0	1	50	0.8	41.7	58.8	53	5	0.204	52.0
56	56.0	3	53	2.5	44.2	57.1	52	5	0.142	51.4
55	55.0	2	55	1.7	45.8	55.0	52	5	0.081	50.8
54	54.0	4	59	3.3	49.2	52.5	51	5	0.019	50.2
53	53.0	3	62	2.5	51.7	49.6	50	5	-0.043	49.6
52	52.0	1	63	0.8	52.5	47.9	50	5	-0.104	49.0
51	51.0	4	67	3.3	55.8	45.8	49	5	-0.166	48.3
50	50.0	1	68	0.8	56.7	43.8	49	5	-0.228	47.7
49	49.0	1	69	0.8	57.5	42.9	49	5	-0.290	47.1
48	48.0	1	70	0.8	58.3	42.1	49	5	-0.351	46.5
47	47.0	5	75	4.2	62.5	39.6	48	6	-0.413	45.9
46	46.0	3	78	2.5	65.0	36.3	47	6	-0.475	45.3
45	45.0	4	82	3.3	68.3	33.3	46	6	-0.536	44.6
44	44.0	2	84	1.7	70.0	30.8	45	6	-0.598	44.0
43	43.0	5	89	4.2	74.2	27.9	45	6	-0.660	43.4
42	42.0	1	90	0.8	75.0	25.4	44	6	-0.722	42.8

HISTOGRAM OF SCORE DISTRIBUTION



ITEM ANALYSIS

ITEM NO.	RESP-ONSE	NUMBER RESPONDING			DIFFICULTY INDICE				CORRELATION COEFFICIENTS					MEAN CRITERION	
		UPPER	MIDDLE	LOWER	UPPER	LOWER	TOTAL	DELTA	DISC.INDEX	BISERIAL RBIS	t	POINT-BISERIAL RPB	t	SCORE	T-SC
1	1	8	18	13	0.250	0.406	0.325	14.86	-0.156	-0.137	-1.50	-0.108	-1.18	51.2	48.45
	2	2	4	5	0.063	0.156	0.092	18.36	-0.094	-0.198	-2.20	-0.128	-1.40	47.2	45.98
	* 3	22	32	4	0.688	0.125	0.483	13.21	0.563	0.463	5.67	0.369	4.32	59.9	53.82
	4	0	2	10	0.000	0.313	0.100	18.17	-0.313	-0.494	-6.17	-0.325	-3.73	37.9	40.26
	ERROR	0	0	0	0.000	0.000	0.000	0.00	0.000	0.000	0.00	0.000	0.00	0.0	0.00
2	1	10	22	9	0.313	0.281	0.342	14.67	0.031	0.019	0.20	0.015	0.16	54.0	50.21
	2	6	19	15	0.188	0.469	0.333	14.76	-0.281	-0.333	-3.84	-0.266	-3.00	47.6	46.24
	3	3	3	3	0.094	0.094	0.075	18.80	0.000	0.100	1.09	0.062	0.68	57.2	52.18
	* 4	13	12	5	0.406	0.156	0.250	15.74	0.250	0.305	3.48	0.235	2.63	60.3	54.08
	ERROR	0	0	0	0.000	0.000	0.000	0.00	0.000	0.000	0.00	0.000	0.00	0.0	0.00
3	1	4	5	4	0.125	0.125	0.108	17.98	0.000	0.020	0.21	0.013	0.14	54.3	50.38
	2	3	4	2	0.094	0.063	0.075	18.80	0.031	-0.060	-0.66	-0.038	-0.41	51.6	48.68
	3	20	38	25	0.625	0.781	0.692	10.96	-0.156	-0.172	-1.90	-0.134	-1.47	52.2	49.10
	* 4	5	9	1	0.156	0.031	0.125	17.64	0.125	0.307	3.50	0.205	2.27	62.5	55.42
	ERROR	0	0	0	0.000	0.000	0.000	0.00	0.000	0.000	0.00	0.000	0.00	0.0	0.00
4	* 1	15	25	3	0.469	0.094	0.358	14.49	0.375	0.413	4.92	0.329	3.78	60.8	54.40
	2	0	8	5	0.000	0.156	0.108	17.98	-0.156	-0.202	-2.24	-0.136	-1.49	47.4	46.11
	3	2	3	2	0.063	0.063	0.058	19.31	0.000	-0.007	-0.08	-0.004	-0.04	53.4	49.84
	4	15	20	22	0.469	0.688	0.475	13.29	-0.219	-0.287	-3.25	-0.229	-2.56	49.8	47.59
	ERROR	0	0	0	0.000	0.000	0.000	0.00	0.000	0.000	0.00	0.000	0.00	0.0	0.00
5	1	0	1	1	0.000	0.031	0.017	21.55	-0.031	-0.206	-2.29	-0.082	-0.89	43.5	43.71
	2	2	6	8	0.063	0.250	0.133	17.48	-0.188	-0.270	-3.05	-0.185	-2.04	46.1	45.29
	* 3	29	37	18	0.906	0.563	0.700	10.86	0.344	0.327	3.76	0.257	2.89	56.4	51.68
	4	1	11	5	0.031	0.156	0.142	17.81	-0.125	-0.187	-2.06	-0.112	-1.22	49.2	47.25
	ERROR	0	1	0	0.000	0.000	0.008	22.61	0.000	-0.207	-2.30	-0.060	-0.66	43.0	43.40
6	1	1	1	3	0.031	0.094	0.042	19.96	-0.063	-0.164	-1.80	-0.086	-0.94	47.0	45.87
	2	1	7	7	0.031	0.219	0.125	17.64	-0.188	-0.376	-4.41	-0.251	-2.82	42.9	43.36
	* 3	29	36	6	0.906	0.188	0.592	12.03	0.719	0.730	11.61	0.586	7.85	61.6	54.87
	4	1	12	16	0.031	0.500	0.242	15.84	-0.469	-0.579	-7.71	-0.439	-5.30	41.1	42.23
	ERROR	0	0	0	0.000	0.000	0.000	0.00	0.000	0.000	0.00	0.000	0.00	0.0	0.00
7	1	2	2	3	0.063	0.094	0.058	19.31	-0.031	-0.170	-1.88	-0.096	-1.05	47.4	46.13
	2	0	10	8	0.000	0.250	0.150	17.18	-0.250	-0.330	-3.79	-0.237	-2.65	44.6	44.36
	* 3	28	39	15	0.875	0.469	0.683	11.05	0.406	0.415	4.95	0.328	3.77	57.3	52.23
	4	2	5	6	0.063	0.188	0.108	17.98	-0.125	-0.217	-2.41	-0.146	-1.60	46.9	45.82
	ERROR	0	0	0	0.000	0.000	0.000	0.00	0.000	0.000	0.00	0.000	0.00	0.0	0.00
8	1	4	12	5	0.125	0.156	0.175	16.78	-0.031	-0.041	-0.45	-0.030	-0.33	52.6	49.34
	* 2	16	15	5	0.500	0.156	0.300	15.14	0.344	0.392	4.62	0.309	3.53	61.3	54.72
	3	8	11	13	0.250	0.406	0.267	15.53	-0.156	-0.159	-1.75	-0.123	-1.35	50.4	47.95
	4	4	18	9	0.125	0.281	0.258	15.63	-0.156	-0.225	-2.51	-0.172	-1.90	49.0	47.08
	ERROR	0	0	0	0.000	0.000	0.000	0.00	0.000	0.000	0.00	0.000	0.00	0.0	0.00

TEST SUMMARY**TEST STATISTICS**

	Mean	Min	Median*	Max	Std Devn	Var
Test Scores	53.692	22.000	57.500	93.000	16.202	262.501
Diff. Index	0.537	0.092	0.483	0.875	0.505	0.255
Delta	12.579	8.187	13.275	18.362	1.991	3.963
Disc. Index	0.403	-0.156	0.281	0.719	0.188	0.035
Biserial (RBIS)	0.441	-0.270	0.258	0.787	0.191	0.037
Point-Biserial (RPB)	0.343	-0.174	0.214	0.602	0.147	0.022

KUDER-RICHARDSON RELIABILITY STATISTICS

KR20 = 0.926 SEM20 = 4.412

KR21 = 0.914 SEM21 = 4.740

CRONBACH ALPHA RELIABILITY STATISTICS

ALPHA = 0.926 SEM-ALP = 4.412

SPLIT-HALF RELIABILITY STATISTICS

RTT = 0.925 SEMTT = 4.449

* Approximate medians,
if the distributions are not normal.

DISTRIBUTION OF DISCRIMINATIONS

Plotted over 20 equal intervals of 0.044

Min	Median*																		Max	
-0.157	-0.069	0.019	0.106	0.194	0.281	0.369	0.456	0.544	0.631	0.720										
	-0.113	-0.025	0.063	0.150	0.238	0.325	0.413	0.500	0.588	0.675										
51	0	0	0	18	59	3	9	39	2	11	5	4	37	13	14	1	22	30	6	
0	0	0	0	83	0	12	16	92	19	27	8	7	42	26	31	10	29	41	52	
0	0	0	0	0	0	77	20	0	25	44	17	23	71	33	50	24	34	97	63	
0	0	0	0	0	0	0	21	0	28	46	43	32	75	40	53	60	35	0	66	
0	0	0	0	0	0	0	65	0	78	80	70	36	87	54	56	88	61	0	72	
0	0	0	0	0	0	0	90	0	96	98	76	38	89	57	62	0	67	0	73	
0	0	0	0	0	0	0	93	0	99	0	0	45	91	74	64	0	69	0	84	
0	0	0	0	0	0	0	94	0	0	0	0	48	0	81	0	0	79	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	49	0	100	0	0	82	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	55	0	0	0	0	95	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	58	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	68	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	85	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	86	0	0	0	0	0	0	0	
Col 1	0	0	0	2	1	3	8	2	7	6	6	14	7	9	7	5	10	3	7	
Totals																				

** If any index is zero,
its item no. is not plotted.

ภาคผนวก ข. ตัวอย่างผลการวิเคราะห์ข้อทดสอบแบบราช (Rasch Model)

RASCH 47

This is ONE-PARAMETER TEST ITEM ANALYSIS PROGRAM

(Rasch Model), Version 8 (2547)

to analyze M/C item.

It was originally written for DOS by

Dr. Suphat Sukamolson, Ph.D. *

(Educational Measurement and Evaluation),

and updated for Windows by

Mr. Komson Eiamjamrus, M.Sc.

LANGUAGE INSTITUTE

CHULALONGKORN UNIVERSITY, 2004

ALL RIGHTS RESERVED

[* Sukamolson, S. *Computerized Item Analyses and Grading*. BKK: Wittayaphat, 2542.]

PROCEDURE IS UCON

DIFFICULTY SCALE FACTOR 1.14 ABILITY SCALE FACTOR 1.20

NUMBER OF ITERATIONS = 2

SEQUENCE NUMBER	ITEM NAME	ITEM DIFFICULTY	STANDARD ERROR	LAST DIFF CHANGE	PROX DIFF	FIRST CYCLE
1	1	0.704	0.140	0.002	0.740	0.702
2	2	0.704	0.140	0.002	0.740	0.702
3	3	1.907	0.169	0.002	1.999	1.905
4	4	-1.330	0.183	-0.004	-1.400	-1.327
5	5	-1.621	0.201	-0.004	-1.716	-1.617
6	6	0.512	0.139	0.002	0.542	0.511
7	7	-1.083	0.171	-0.003	-1.133	-1.081
8	8	1.315	0.149	0.002	1.376	1.313
9	9	-0.311	0.146	0.000	-0.314	-0.310
10	10	-1.790	0.213	-0.004	-1.902	-1.786
11	11	1.061	0.144	0.002	1.110	1.058
12	12	0.360	0.139	0.001	0.384	0.359
13	13	0.360	0.139	0.001	0.384	0.359
14	14	-1.469	0.191	-0.003	-1.550	-1.465
15	15	0.284	0.139	0.001	0.305	0.283
16	16	0.246	0.139	0.001	0.266	0.245
17	17	0.781	0.140	0.002	0.821	0.779
18	18	-1.330	0.183	-0.004	-1.400	-1.327
19	19	0.474	0.139	0.002	0.503	0.473
20	20	0.781	0.140	0.002	0.821	0.779
21	21	1.359	0.150	0.002	1.422	1.357
22	22	-0.105	0.142	0.000	-0.100	-0.106
23	23	-1.265	0.179	-0.003	-1.329	-1.262
24	24	-0.439	0.149	-0.001	-0.449	-0.439
25	25	0.436	0.139	0.001	0.463	0.435
26	26	0.743	0.140	0.002	0.780	0.741
27	27	-0.439	0.149	-0.001	-0.449	-0.439
28	28	0.665	0.139	0.002	0.700	0.664
29	29	-1.790	0.213	-0.004	-1.902	-1.786
30	30	-0.764	0.158	-0.002	-0.792	-0.762
31	31	-1.142	0.173	-0.003	-1.196	-1.139
32	32	-0.667	0.155	-0.001	-0.690	-0.666
33	33	-0.396	0.148	-0.001	-0.404	-0.395
34	34	-1.202	0.176	-0.003	-1.262	-1.199
35	35	-0.667	0.155	-0.001	-0.690	-0.666
36	36	-0.227	0.144	0.000	-0.227	-0.227
37	37	0.014	0.141	0.000	0.024	0.013
38	38	1.315	0.149	0.002	1.376	1.313
39	39	-0.227	0.144	0.000	-0.227	-0.227
40	40	-0.439	0.149	-0.001	-0.449	-0.439
41	41	0.053	0.140	0.001	0.065	0.052
42	42	1.315	0.149	0.002	1.376	1.313

SEQUENCE NUMBER	ITEM NAME	ITEM DIFFICULTY	STANDARD ERROR	LAST DIFF CHANGE	PROX DIFF	FIRST CYCLE
43	43	0.014	0.141	0.000	0.024	0.013
44	44	0.474	0.139	0.002	0.503	0.473
45	45	-0.667	0.155	-0.001	-0.690	-0.666
46	46	0.627	0.139	0.002	0.661	0.625
47	47	1.143	0.145	0.002	1.197	1.141
48	48	0.550	0.139	0.002	0.581	0.549
49	49	0.781	0.140	0.002	0.821	0.779
50	50	-2.339	0.263	-0.005	-2.511	-2.334
51	51	0.436	0.139	0.001	0.463	0.435
52	52	0.130	0.140	0.001	0.146	0.130
53	53	-0.227	0.144	0.000	-0.227	-0.227
54	54	1.271	0.148	0.002	1.330	1.269
55	55	1.359	0.150	0.002	1.422	1.357
56	56	-0.396	0.148	-0.001	-0.404	-0.395
57	57	0.589	0.139	0.002	0.621	0.587
58	58	0.053	0.140	0.001	0.065	0.052
59	59	-0.483	0.150	-0.001	-0.496	-0.483
60	60	1.020	0.143	0.002	1.068	1.018
61	61	-0.227	0.144	0.000	-0.227	-0.227
62	62	-1.265	0.179	-0.003	-1.329	-1.262
63	63	-0.227	0.144	0.000	-0.227	-0.227
64	64	1.271	0.148	0.002	1.330	1.269
65	65	-0.311	0.146	0.000	-0.314	-0.310
66	66	0.979	0.143	0.002	1.026	0.977
67	67	1.450	0.153	0.002	1.517	1.448
68	68	1.143	0.145	0.002	1.197	1.141
69	69	0.053	0.140	0.001	0.065	0.052
70	70	0.860	0.141	0.002	0.902	0.858
71	71	0.512	0.139	0.002	0.542	0.511
72	72	1.692	0.160	0.002	1.772	1.690
73	73	-0.066	0.142	0.000	-0.058	-0.066
74	74	1.692	0.160	0.002	1.772	1.690
75	75	0.360	0.139	0.001	0.384	0.359
76	76	2.214	0.184	0.002	2.326	2.212
77	77	-0.667	0.155	-0.001	-0.690	-0.666
78	78	1.744	0.162	0.002	1.826	1.742
79	79	0.704	0.140	0.002	0.740	0.702
80	80	1.061	0.144	0.002	1.110	1.058
81	81	-2.485	0.279	-0.005	-2.674	-2.479
82	82	-2.091	0.238	-0.005	-2.235	-2.086
83	83	0.053	0.140	0.001	0.065	0.052
84	84	0.246	0.139	0.001	0.266	0.245
85	85	-0.396	0.148	-0.001	-0.404	-0.395
86	86	-0.353	0.147	0.000	-0.359	-0.353
87	87	0.743	0.140	0.002	0.780	0.741
88	88	-2.651	0.299	-0.006	-2.861	-2.645
89	89	-0.814	0.160	-0.002	-0.845	-0.812

SEQUENCE NUMBER	ITEM NAME	ITEM DIFFICULTY	STANDARD ERROR	LAST DIFF CHANGE	PROX DIFF	FIRST CYCLE
90	90	-2.209	0.249	-0.005	-2.366	-2.204
91	91	0.474	0.139	0.002	0.503	0.473
92	92	-2.091	0.238	-0.005	-2.235	-2.086
93	93	-0.396	0.148	-0.001	-0.404	-0.395
94	94	0.207	0.139	0.001	0.226	0.206
95	95	-1.330	0.183	-0.004	-1.400	-1.327
96	96	0.589	0.139	0.002	0.621	0.587
97	97	-0.026	0.141	0.000	-0.017	-0.026
98	98	-0.353	0.147	0.000	-0.359	-0.353
99	99	0.092	0.140	0.001	0.106	0.091
100	100	-1.202	0.176	-0.003	-1.262	-1.199

ROOT MEAN SQUARE = 0.002

100 ITEMS CALIBRATED ON 232 PERSONS

232 MEASURABLE PERSONS WITH MEAN ABILITY = 0.44 AND STD. DEV. = 0.66

MAP OF VARIABLE

PERSON STATS	COUNT	RAW SCORE	MEASURE MIDPOINT (S.E.)	ITEM COUNTS	TYPICAL ITEMS (BY NAME)															
+3SD	2	89	2.49 (0.33)																	
		87	2.29 (0.31)	1	76															
	2	85	2.09 (0.30)																	
	4	82	1.89 (0.28)	1	3															
+2SD	6	79	1.69 (0.26)	3	72	74	78													
	10	76	1.49 (0.25)	1	67															
	10	73	1.29 (0.24)	7	8	21	38	42	54	55	64									
+1SD	16	70	1.09 (0.24)	5	11	47	60	68	80											
	12	66	0.89 (0.23)	2	66	70														
	26	62	0.69 (0.23)	10	1	2	17	20	26	28	46	49	79	87						
MEAN	26	58	0.49 (0.22)	10	6	19	25	44	48	51	57	71	91	96						
	36	54	0.29 (0.22)	7	12	13	15	16	75	84	94									
	16	50	0.09 (0.22)	8	37	41	43	52	58	69	83	99								
	20	46	-0.11 (0.22)	3	22	73	97													
-1SD	18	42	-0.31 (0.23)	13	9	33	36	39	53	56	61	63	65	85						
	18	38	-0.51 (0.23)	4	24	27	40	59												
	2	34	-0.71 (0.23)	5	30	32	35	45	77											
-2SD	4	30	-0.91 (0.24)	1	89															
	2	27	-1.11 (0.25)	4	7	31	34	100												
	2	24	-1.31 (0.26)	5	4	18	23	62	95											
-3SD		21	-1.51 (0.27)	1	14															
		19	-1.71 (0.28)	3	5	10	29													
		16	-1.91 (0.30)																	
		14	-2.11 (0.31)	3	82	90	92													
-4SD		12	-2.31 (0.33)	1	50															
		10	-2.51 (0.36)	1	81															
		9	-2.71 (0.37)	1	88															

100 ITEMS CALIBRATED ON 232 PERSONS

232 MEASURABLE PERSONS WITH MEAN ABILITY = 0.44 AND STD. DEV. = 0.66

ITEM CHARACTERISTIC CURVE							DEPARTURE FROM EXPECTED ICC						ITEM FIT STATISTICS							
SEQ	ITEM	1ST	2ND	3RD	4TH	5TH	6TH	1ST	2ND	3RD	4TH	5TH	6TH	* ERR	FIT	T-TESTS	WEIGHTED	DISC	POINT	
NUM	NAME	GROUP	GROUP	GROUP	GROUP	GROUP	GROUP	GROUP	GROUP	GROUP	GROUP	GROUP	GROUP	IMPAC	BETWN	TOTAL	MNSQ	DF	INDX	BISER
1	1	0.45	0.24	0.40	0.58	0.58	0.41	0.23	-0.09	-0.01	0.11	0.01	-0.29	0.08	4.00	1.72	1.18	202	0.16	0.06
2	2	0.35	0.14	0.35	0.42	0.63	0.82	0.13	-0.18	-0.06	-0.05	0.07	0.12	0.00	2.21	-0.71	0.93	202	1.22	0.40
3	3	0.15	0.14	0.15	0.16	0.37	0.29	0.07	0.02	-0.02	-0.05	0.09	-0.12	0.02	0.83	0.42	1.05	138	0.61	0.16
4	4	0.70	0.81	0.85	0.79	0.89	1.00	0.02	0.03	0.01	-0.08	-0.01	0.05	0.01	0.07	0.13	1.01	118	1.00	0.22
5	5	0.45	0.90	0.95	1.00	0.95	1.00	-0.29	0.08	0.07	0.10	0.02	0.04	0.00	3.79	-1.27	0.82	98	2.06	0.49
6	6	0.30	0.48	0.40	0.58	0.32	0.88	0.04	0.11	-0.06	0.06	-0.30	0.15	0.02	3.18	0.44	1.04	205	0.82	0.26
7	7	0.65	0.67	0.90	0.89	0.84	0.88	0.03	-0.07	0.09	0.05	-0.04	-0.05	0.00	0.62	-0.06	0.99	136	0.84	0.24
8	8	0.35	0.10	0.30	0.11	0.47	0.59	0.22	-0.11	0.03	-0.22	0.06	0.03	0.02	3.93	0.43	1.04	177	0.72	0.20
9	9	0.35	0.62	0.75	0.79	0.68	0.82	-0.09	0.05	0.09	0.08	-0.10	-0.04	0.00	0.77	0.03	1.00	185	0.95	0.29
10	10	0.60	0.90	0.95	0.95	0.95	1.00	-0.17	0.05	0.06	0.03	0.01	0.04	0.00	1.56	-0.43	0.93	87	1.69	0.33
11	11	0.25	0.00	0.35	0.37	0.79	0.47	0.08	-0.25	0.02	-0.31	0.31	-0.15	0.00	4.51	-0.20	0.98	190	1.10	0.35
12	12	0.45	0.29	0.40	0.53	0.68	0.82	0.16	-0.12	-0.09	-0.03	0.04	0.06	0.01	1.48	-0.15	1.01	205	0.95	0.29
13	13	0.50	0.33	0.45	0.58	0.58	0.71	0.21	-0.07	-0.04	0.02	-0.07	-0.06	0.05	1.78	0.94	1.09	205	0.49	0.18
14	14	0.70	0.67	1.00	0.79	1.00	1.00	-0.01	-0.14	0.14	-0.10	0.08	0.05	0.00	3.05	-0.36	0.95	108	1.43	0.32
15	15	0.45	0.29	0.35	0.84	0.53	0.82	0.15	-0.14	-0.16	0.27	-0.14	0.05	0.01	3.74	0.22	1.02	204	0.89	0.28
16	16	0.35	0.33	0.70	0.37	0.68	0.88	0.04	-0.10	0.18	-0.21	0.01	0.10	0.00	2.48	-0.30	0.97	204	1.12	0.36
17	17	0.30	0.24	0.35	0.32	0.53	0.88	0.09	-0.07	-0.04	-0.13	-0.02	0.20	0.00	1.90	-0.54	0.94	200	1.21	0.37
18	18	0.45	0.76	0.90	0.95	1.00	1.00	-0.23	-0.02	0.06	0.08	0.09	0.05	0.00	2.84	-0.85	0.89	118	2.02	0.44
19	19	0.45	0.38	0.50	0.42	0.58	0.65	0.19	0.00	0.03	-0.11	-0.04	-0.10	0.04	1.66	0.93	1.09	205	0.38	0.19
20	20	0.25	0.24	0.45	0.37	0.53	0.76	0.04	-0.07	0.06	-0.08	-0.02	0.08	0.00	-0.09	-0.05	0.99	200	1.09	0.31
21	21	0.25	0.48	0.20	0.21	0.32	0.35	0.12	0.28	-0.07	-0.10	-0.09	-0.20	0.08	4.53	1.47	1.16	175	-0.03	0.04
22	22	0.50	0.48	0.50	0.74	0.68	0.88	0.11	-0.04	-0.11	0.07	-0.06	0.04	0.02	0.66	0.45	1.04	195	0.88	0.25
23	23	0.80	0.76	0.75	0.84	0.84	1.00	0.14	-0.01	-0.08	-0.02	-0.06	0.06	0.02	1.27	0.42	1.05	123	0.71	0.17
24	24	0.20	0.67	0.75	0.79	0.89	0.88	-0.27	0.07	0.06	0.05	0.09	0.00	0.00	2.45	-1.21	0.87	179	1.62	0.47
25	25	0.40	0.48	0.25	0.68	0.42	0.82	0.13	0.09	-0.23	0.15	-0.21	0.07	0.04	3.52	0.82	1.08	205	0.68	0.20
26	26	0.30	0.43	0.35	0.53	0.47	0.53	0.08	0.11	-0.05	0.07	-0.08	-0.16	0.05	1.56	0.96	1.10	201	0.38	0.19
27	27	0.25	0.71	0.65	0.79	0.79	1.00	-0.22	0.11	-0.04	0.05	-0.01	0.12	0.00	2.42	-0.77	0.92	179	1.61	0.41
28	28	0.25	0.57	0.45	0.53	0.26	0.65	0.02	0.24	0.03	0.05	-0.31	-0.06	0.06	3.81	1.24	1.12	203	0.33	0.14
29	29	0.65	0.81	1.00	1.00	0.95	0.94	-0.12	-0.04	0.11	0.09	0.01	-0.02	0.00	1.95	-0.30	0.95	87	1.47	0.29
30	30	0.60	0.76	0.65	0.74	0.95	0.82	0.05	0.09	-0.10	-0.06	0.10	-0.08	0.02	1.57	0.33	1.03	158	0.71	0.22
31	31	0.60	0.81	0.85	0.74	0.95	0.94	-0.04	0.06	0.04	-0.11	0.06	0.01	0.00	0.59	-0.33	0.96	131	1.06	0.30
32	32	0.35	0.71	0.60	0.79	1.00	1.00	-0.17	0.06	-0.13	0.01	0.16	0.10	0.00	3.07	-1.19	0.87	165	1.77	0.48
33	33	0.50	0.67	0.65	0.68	0.74	0.88	0.04	0.08	-0.03	-0.04	-0.06	0.01	0.01	-0.70	0.29	1.03	181	0.78	0.26
34	34	0.30	0.81	0.85	1.00	1.00	1.00	-0.35	0.05	0.03	0.14	0.10	0.06	0.00	4.61	-1.45	0.82	127	2.27	0.53
35	35	0.25	0.67	0.75	0.89	0.95	0.94	-0.27	0.01	0.02	0.12	0.11	0.04	0.00	2.94	-1.53	0.84	165	1.86	0.52
36	36	0.30	0.48	0.65	0.79	0.84	0.88	-0.12	-0.07	0.01	0.10	0.08	0.03	0.00	0.60	-1.08	0.89	189	1.47	0.46
37	37	0.40	0.62	0.40	0.74	0.79	0.65	0.04	0.13	-0.18	0.10	0.07	-0.17	0.04	2.75	0.77	1.08	199	0.59	0.20
38	38	0.05	0.19	0.35	0.37	0.26	0.71	-0.08	-0.02	0.08	0.04	-0.15	0.14	0.00	1.57	-0.46	0.95	177	1.32	0.37
39	39	0.35	0.52	0.70	0.63	0.79	0.94	-0.07	-0.02	0.06	-0.06	0.02	0.09	0.00	0.01	-0.72	0.93	189	1.33	0.41
40	40	0.55	0.71	0.45	0.63	0.95	0.88	0.08	0.11	-0.24	-0.10	0.14	0.00	0.02	3.10	0.43	1.04	179	0.89	0.24
41	41	0.40	0.52	0.50	0.58	0.79	0.76	0.05	0.05	-0.07	-0.05	0.08	-0.05	0.01	-0.26	0.29	1.03	200	0.84	0.27
42	42	0.10	0.24	0.35	0.21	0.37	0.65	-0.03	0.03	0.08	-0.11	-0.05	0.09	0.00	0.36	-0.21	0.97	177	1.11	0.32
43	43	0.25	0.48	0.75	0.42	0.74	1.00	-0.11	-0.01	0.17	-0.22	0.02	0.18	0.00	3.27	-0.71	0.93	199	1.49	0.42
44	44	0.15	0.43	0.55	0.58	0.58	0.71	-0.11	0.05	0.08	0.05	-0.04	-0.04	0.00	0.30	-0.19	0.98	205	1.06	0.35
45	45	0.50	0.67	0.90	0.84	0.68	0.82	-0.02	0.01	0.17	0.07	-0.15	-0.08	0.01	2.38	0.20	1.02	165	0.68	0.22
46	46	0.20	0.33	0.40	0.37	0.68	0.82	-0.04	-0.01	-0.03	-0.12	0.10	0.11	0.00	0.57	-0.28	0.97	204	1.35	0.36
47	47	0.35	0.33	0.35	0.26	0.37	0.41	0.19	0.10	0.04	-0.10	-0.09	-0.19	0.08	3.26	1.56	1.17	186	0.03	0.07
48	48	0.30	0.43	0.35	0.47	0.53	0.82	0.05	0.07	-0.10	-0.33	-0.08	0.09	0.01	0.40	0.19	1.02	205	0.94	0.29
49	49	0.30	0.57	0.35	0.26	0.37	0.71	0.09	0.26	-0.04	-0.19	-0.18	0.02	0.04	3.73	0.85	1.08	200	0.40	0.19
50	50	0.70	1.00	0.90	1.00	1.00	1.00	-0.15	0.09	-0.04	0.05	0.04	0.02	0.00	2.54	-0.57	0.89	57	1.87	0.35

RASCH 47

This is ONE-PARAMETER TEST ITEM ANALYSIS PROGRAM

(Rasch Model), Version 8 (2547)

to analyze M/C item.

It was originally written for DOS by

Dr. Suphat Sukamolson, Ph.D. *

(Educational Measurement and Evaluation),

and updated for Windows by

Mr. Komson Eiamjamrus, M.Sc.

LANGUAGE INSTITUTE

CHULALONGKORN UNIVERSITY, 2004

ALL RIGHTS RESERVED

[* Sukamolson, S. *Computerized Item Analyses and Grading*. BKK: Wittayaphat, 2542.]

PROCEDURE IS UCON

DIFFICULTY SCALE FACTOR 1.20 ABILITY SCALE FACTOR 1.21

NUMBER OF ITERATIONS = 2

SEQUENCE NUMBER	ITEM NAME	ITEM DIFFICULTY	STANDARD ERROR	LAST DIFF CHANGE	PROX DIFF	FIRST CYCLE
1	1	-0.452	0.143	-0.003	-0.460	-0.449
2	2	1.763	0.164	0.011	1.842	1.752
3	3	-0.331	0.142	-0.003	-0.333	-0.329
4	4	0.818	0.141	0.006	0.856	0.812
5	5	0.585	0.139	0.004	0.615	0.580
6	6	0.206	0.138	0.001	0.225	0.205
7	7	-0.748	0.149	-0.006	-0.773	-0.743
8	8	-0.534	0.145	-0.004	-0.546	-0.530
9	9	-0.452	0.143	-0.003	-0.460	-0.449
10	10	-0.452	0.143	-0.003	-0.460	-0.449
11	11	0.740	0.141	0.005	0.775	0.734
12	12	0.898	0.143	0.007	0.938	0.892
13	13	-0.576	0.146	-0.004	-0.590	-0.572
14	14	1.461	0.154	0.010	1.524	1.451
15	15	0.319	0.138	0.002	0.342	0.317
16	16	0.319	0.138	0.002	0.342	0.317
17	17	0.546	0.139	0.004	0.576	0.542
18	18	0.546	0.139	0.004	0.576	0.542
19	19	0.319	0.138	0.002	0.342	0.317
20	20	0.779	0.141	0.006	0.815	0.773
21	21	-0.618	0.146	-0.005	-0.635	-0.614
22	22	1.509	0.156	0.010	1.574	1.499
23	23	0.700	0.140	0.005	0.734	0.695
24	24	-0.371	0.142	-0.003	-0.375	-0.368
25	25	-0.661	0.147	-0.005	-0.680	-0.656
26	26	-1.457	0.174	-0.009	-1.544	-1.448
27	27	-1.518	0.177	-0.010	-1.612	-1.509
28	28	1.104	0.146	0.008	1.152	1.097
29	29	0.093	0.138	0.001	0.108	0.092
30	30	-0.253	0.141	-0.002	-0.250	-0.251
31	31	0.898	0.143	0.007	0.938	0.892
32	32	1.278	0.150	0.009	1.332	1.269
33	33	-1.026	0.157	-0.007	-1.072	-1.019
34	34	0.206	0.138	0.001	0.225	0.205
35	35	-0.331	0.142	-0.003	-0.333	-0.329
36	36	-0.793	0.150	-0.006	-0.821	-0.787
37	37	-1.126	0.160	-0.008	-1.181	-1.119
38	38	-1.581	0.180	-0.010	-1.682	-1.572
39	39	-1.715	0.187	-0.010	-1.832	-1.705
40	40	-0.931	0.154	-0.007	-0.969	-0.924
41	41	-1.076	0.159	-0.007	-1.126	-1.069
42	42	-0.704	0.148	-0.005	-0.727	-0.699

SEQUENCE NUMBER	ITEM NAME	ITEM DIFFICULTY	STANDARD ERROR	LAST DIFF CHANGE	PROX DIFF	FIRST CYCLE
43	43	-0.021	0.139	0.000	-0.010	-0.021
44	44	0.055	0.138	0.000	0.069	0.055
45	45	1.368	0.152	0.009	1.426	1.359
46	46	1.873	0.169	0.012	1.959	1.862
47	47	-0.748	0.149	-0.006	-0.773	-0.743
48	48	-0.097	0.139	-0.001	-0.089	-0.097
49	49	-0.793	0.150	-0.006	-0.821	-0.787
50	50	2.724	0.216	0.014	2.884	2.710
51	51	0.508	0.139	0.004	0.536	0.504
52	52	-0.704	0.148	-0.005	-0.727	-0.699
53	53	-0.371	0.142	-0.003	-0.375	-0.368
54	54	0.093	0.138	0.001	0.108	0.092
55	55	-1.126	0.160	-0.008	-1.181	-1.119
56	56	-0.331	0.142	-0.003	-0.333	-0.329
57	57	-0.576	0.146	-0.004	-0.590	-0.572
58	58	2.247	0.186	0.013	2.361	2.234
59	59	-0.136	0.140	-0.001	-0.129	-0.135
60	60	-0.838	0.151	-0.006	-0.869	-0.832
61	61	1.658	0.161	0.011	1.731	1.648
62	62	0.131	0.138	0.001	0.147	0.130
63	63	-0.059	0.139	-0.001	-0.049	-0.059
64	64	1.608	0.159	0.011	1.678	1.597
65	65	0.662	0.140	0.005	0.694	0.657
66	66	-0.136	0.140	-0.001	-0.129	-0.135
67	67	0.470	0.138	0.003	0.497	0.467
68	68	0.980	0.144	0.007	1.022	0.972
69	69	1.710	0.162	0.011	1.786	1.699
70	70	0.093	0.138	0.001	0.108	0.092
71	71	-0.021	0.139	0.000	-0.010	-0.021
72	72	1.931	0.171	0.012	2.021	1.919
73	73	1.763	0.164	0.011	1.842	1.752
74	74	0.546	0.139	0.004	0.576	0.542
75	75	0.623	0.139	0.005	0.655	0.618
76	76	0.206	0.138	0.001	0.225	0.205
77	77	1.190	0.148	0.008	1.241	1.182
78	78	1.558	0.157	0.010	1.625	1.548
79	79	1.414	0.153	0.010	1.475	1.405
80	80	1.558	0.157	0.010	1.625	1.548
81	81	-1.178	0.162	-0.008	-1.237	-1.170
82	82	-2.412	0.238	-0.013	-2.627	-2.399
83	83	-0.793	0.150	-0.006	-0.821	-0.787
84	84	-1.581	0.180	-0.010	-1.682	-1.572
85	85	-2.412	0.238	-0.013	-2.627	-2.399
86	86	-2.022	0.206	-0.011	-2.179	-2.011
87	87	-1.341	0.169	-0.009	-1.415	-1.332
88	88	-1.398	0.171	-0.009	-1.478	-1.389
89	89	-1.178	0.162	-0.008	-1.237	-1.170

SEQUENCE NUMBER	ITEM NAME	ITEM DIFFICULTY	STANDARD ERROR	LAST DIFF CHANGE	PROX DIFF	FIRST CYCLE
90	90	-0.021	0.139	0.000	-0.010	-0.021
91	91	0.017	0.138	0.000	0.030	0.017
92	92	-1.231	0.164	-0.008	-1.295	-1.223
93	93	-0.931	0.154	-0.007	-0.969	-0.924
94	94	1.509	0.156	0.010	1.574	1.499
95	95	0.546	0.139	0.004	0.576	0.542
96	96	-0.748	0.149	-0.006	-0.773	-0.743
97	97	0.017	0.138	0.000	0.030	0.017
98	98	-1.715	0.187	-0.010	-1.832	-1.705
99	99	0.131	0.138	0.001	0.147	0.130
100	100	-1.647	0.183	-0.010	-1.755	-1.637

ROOT MEAN SQUARE = 0.007

100 ITEMS CALIBRATED ON 248 PERSONS

248 MEASURABLE PERSONS WITH MEAN ABILITY = 0.29 AND STD. DEV. = 0.85

COMPLETE SCORE EQUIVALENCE TABLE

RAW SCORE	COUNT	LOG ABILITY	STANDARD ERRORS	TEST CHARACTERISTIC CURVE
11	0	-2.48	0.34	*
10	0	-2.59	0.35	*
9	0	-2.72	0.37	*
8	0	-2.86	0.38	*
7	0	-3.01	0.41	*
6	0	-3.19	0.44	*
5	0	-3.39	0.47	*
4	0	-3.64	0.52	*
3	0	-3.94	0.60	*
2	0	-4.37	0.72	*
1	0	-5.08	1.01	*

PERSON SEPARABILITY INDEX 0.93 (EQUIVALENT TO KR 20)

100 ITEMS CALIBRATED ON 248 PERSONS

248 MEASURABLE PERSONS WITH MEAN ABILITY = 0.29 AND STD. DEV. = 0.85

MAP OF VARIABLE

PERSON STATS	COUNT	RAW SCORE	MEASURE MIDPOINT (S.E.)	ITEM COUNTS	TYPICAL ITEMS (BY NAME)															
+3SD	2	92	2.90 (0.39)																	
	2	90	2.70 (0.35)	1	50															
		89	2.50 (0.34)																	
		87	2.30 (0.32)	1	58															
+2SD	4	84	2.10 (0.29)																	
	4	82	1.90 (0.28)	2	46	72														
	2	79	1.70 (0.27)	5	2	61	64	69	73											
	16	76	1.50 (0.26)	6	14	22	78	79	80	94										
	12	73	1.30 (0.25)	2	32	45														
+1SD	12	70	1.10 (0.24)	2	28	77														
	22	66	0.90 (0.24)	4	4	12	31	68												
	12	63	0.70 (0.23)	5	11	20	23	65	75											
	22	59	0.50 (0.23)	7	5	17	18	51	67	74	95									
MEAN	16	55	0.30 (0.23)	6	6	15	16	19	34	76										
	18	51	0.10 (0.22)	8	29	44	54	62	70	91	97	99								
	26	47	-0.10 (0.22)	7	43	48	59	63	66	71	90									
	18	43	-0.30 (0.23)	6	3	24	30	35	53	56										
-1SD	14	39	-0.50 (0.23)	6	1	8	9	10	13	57										
	16	35	-0.70 (0.23)	10	7	21	25	36	42	47	49	52	83	96						
	16	31	-0.90 (0.24)	3	40	60	93													
	10	28	-1.10 (0.24)	6	33	37	41	55	81	89										
	4	25	-1.30 (0.25)	3	87	88	92													
-2SD		22	-1.50 (0.26)	4	26	27	38	84												
		19	-1.70 (0.27)	3	39	98	100													
		16	-1.90 (0.29)																	
		14	-2.10 (0.31)	1	86															
-3SD		12	-2.30 (0.33)																	
		10	-2.50 (0.35)	2	82	85														

100 ITEMS CALIBRATED ON 248 PERSONS

248 MEASURABLE PERSONS WITH MEAN ABILITY = 0.29 AND STD. DEV. = 0.85

ITEM CHARACTERISTIC CURVE							DEPARTURE FROM EXPECTED ICC						ITEM FIT STATISTICS							
SEQ	ITEM	1ST	2ND	3RD	4TH	5TH	6TH	1ST	2ND	3RD	4TH	5TH	6TH	* ERR	FIT	T-TESTS	WEIGHTED	DISC	POINT	
NUM	NAME	GROUP	GROUP	GROUP	GROUP	GROUP	GROUP	GROUP	GROUP	GROUP	GROUP	GROUP	GROUP	IMPAC	BETWN	TOTAL	MNSQ	DF	INDX	BISER
1	1	0.14	0.67	0.64	0.65	1.00	0.81	-0.24	0.15	0.03	-0.06	0.20	-0.08	0.00	3.92	-0.95	0.90	192	1.29	0.46
2	2	0.10	0.06	0.18	0.30	0.36	0.29	0.03	-0.05	0.04	0.09	0.06	-0.19	0.06	1.63	0.99	1.12	146	0.67	0.23
3	3	0.29	0.33	0.73	0.70	0.82	0.86	-0.07	-0.15	0.15	0.02	0.04	-0.02	0.00	1.15	-0.83	0.92	197	1.18	0.46
4	4	0.10	0.22	0.32	0.40	0.55	0.71	-0.05	0.00	0.01	-0.01	0.02	0.02	0.00	-1.58	-0.45	0.95	197	1.14	0.43
5	5	0.24	0.28	0.59	0.45	0.32	0.71	0.06	0.01	0.24	-0.01	-0.26	-0.02	0.07	3.53	1.38	1.14	204	0.55	0.24
6	6	0.48	0.44	0.36	0.45	0.64	0.71	0.23	0.09	-0.08	-0.11	-0.03	-0.09	0.08	2.89	1.63	1.17	207	0.40	0.20
7	7	0.57	0.50	0.68	0.75	0.77	0.95	0.12	-0.08	0.01	-0.02	-0.07	0.04	0.02	0.46	0.46	1.05	177	0.87	0.30
8	8	0.14	0.56	0.73	0.85	0.86	0.86	-0.26	0.02	0.10	0.13	0.05	-0.04	0.00	2.79	-1.52	0.85	188	1.37	0.51
9	9	0.14	0.50	0.68	0.80	0.82	0.95	-0.24	-0.01	0.07	0.09	0.02	0.07	0.00	2.31	-1.32	0.87	192	1.53	0.50
10	10	0.29	0.39	0.59	0.85	0.86	0.90	-0.10	-0.12	-0.02	0.14	0.07	0.02	0.00	1.26	-1.08	0.89	192	1.39	0.48
11	11	0.52	0.39	0.36	0.30	0.36	0.48	0.36	0.15	0.04	-0.13	-0.18	-0.23	0.17	6.68	3.33	1.37	200	-0.28	-0.03
12	12	0.24	0.06	0.23	0.40	0.55	0.71	0.10	-0.16	-0.06	0.01	0.04	0.04	0.00	1.48	-0.62	0.94	194	1.10	0.44
13	13	0.48	0.56	0.45	0.75	0.86	0.95	0.06	0.01	-0.18	0.02	0.05	0.06	0.00	1.27	-0.07	0.99	186	1.11	0.38
14	14	0.19	0.22	0.36	0.20	0.36	0.24	0.11	0.09	0.18	-0.07	-0.01	-0.31	0.13	4.56	2.30	1.27	165	0.07	0.04
15	15	0.52	0.33	0.41	0.40	0.41	0.86	0.30	0.01	-0.01	-0.13	-0.24	0.07	0.09	4.70	1.79	1.18	207	0.44	0.18
16	16	0.24	0.22	0.45	0.45	0.59	0.95	0.01	-0.10	0.04	-0.08	-0.05	0.17	0.00	1.62	-0.66	0.93	207	1.28	0.45
17	17	0.29	0.33	0.45	0.40	0.50	0.67	0.10	0.06	0.09	-0.07	-0.09	-0.08	0.08	1.15	1.54	1.16	204	0.54	0.22
18	18	0.38	0.39	0.32	0.35	0.50	0.71	0.19	0.11	-0.05	-0.12	-0.09	-0.03	0.07	2.64	1.49	1.15	204	0.48	0.21
19	19	0.14	0.44	0.45	0.55	0.59	0.76	-0.08	0.12	0.04	0.02	-0.05	-0.02	0.01	0.18	0.30	1.03	207	0.94	0.36
20	20	0.19	0.39	0.23	0.20	0.73	0.62	0.04	0.15	-0.08	-0.22	0.19	-0.08	0.01	3.27	0.26	1.02	198	0.84	0.35
21	21	0.24	0.33	0.73	0.80	0.95	1.00	-0.18	-0.22	0.08	0.06	0.13	0.10	0.00	3.55	-1.85	0.82	184	1.78	0.56
22	22	0.24	0.17	0.32	0.15	0.23	0.43	0.16	0.04	0.14	-0.11	-0.13	-0.11	0.07	3.87	1.32	1.15	162	0.29	0.14
23	23	0.19	0.28	0.23	0.25	0.64	0.86	0.03	0.03	-0.10	-0.18	0.08	0.14	0.00	2.08	-0.82	0.92	201	1.26	0.46
24	24	0.33	0.39	0.55	0.60	0.91	1.00	-0.03	-0.10	-0.04	-0.09	0.13	0.12	0.00	2.09	-1.10	0.89	195	1.47	0.50
25	25	0.67	0.50	0.50	0.80	0.73	0.95	0.23	-0.06	-0.16	0.05	-0.10	0.05	0.05	2.95	0.97	1.10	182	0.73	0.24
26	26	0.52	0.78	0.73	0.90	1.00	1.00	-0.10	0.04	-0.08	0.03	0.09	0.05	0.00	1.55	-0.91	0.89	130	1.45	0.43
27	27	0.71	0.61	0.91	0.90	0.86	0.95	0.07	-0.14	0.09	0.02	-0.06	0.00	0.01	1.31	0.14	1.01	126	0.86	0.25
28	28	0.10	0.17	0.23	0.30	0.50	0.67	-0.02	-0.01	-0.02	-0.04	0.05	0.04	0.00	-1.55	-0.47	0.95	185	1.15	0.43
29	29	0.10	0.33	0.45	0.60	0.77	0.95	-0.17	-0.04	-0.02	0.01	0.08	0.13	0.00	2.00	-1.74	0.84	207	1.61	0.56
30	30	0.24	0.33	0.64	0.70	0.77	0.95	-0.10	-0.13	0.08	0.03	0.01	0.09	0.00	1.10	-1.17	0.88	199	1.40	0.50
31	31	0.14	0.17	0.36	0.40	0.41	0.71	0.00	-0.05	0.08	0.01	-0.10	0.04	0.00	-0.22	-0.04	0.99	194	0.99	0.39
32	32	0.10	0.17	0.14	0.50	0.45	0.43	0.00	0.01	-0.08	0.20	0.04	-0.16	0.02	2.17	0.32	1.03	176	0.81	0.33
33	33	0.43	0.39	0.82	0.90	0.95	1.00	-0.02	-0.26	0.09	0.09	0.08	0.07	0.00	3.19	-1.30	0.86	160	1.61	0.50
34	34	0.33	0.17	0.41	0.40	0.82	0.90	0.09	-0.18	-0.04	-0.16	0.15	0.10	0.00	2.83	-0.93	0.91	207	1.29	0.48
35	35	0.38	0.33	0.59	0.70	0.91	0.81	0.03	-0.15	0.01	0.02	0.13	-0.06	0.00	1.31	-0.30	0.97	197	1.07	0.41
36	36	0.52	0.56	0.73	0.75	0.91	0.81	0.06	-0.04	0.04	-0.02	0.06	-0.11	0.02	1.19	0.45	1.04	174	0.76	0.26
37	37	0.48	0.61	0.82	0.75	0.95	1.00	-0.07	-0.06	0.07	-0.08	0.07	0.06	0.00	1.24	-0.74	0.91	153	1.35	0.43
38	38	0.43	0.72	0.91	0.95	1.00	1.00	-0.23	-0.04	0.08	0.07	0.08	0.04	0.00	2.90	-1.19	0.85	122	1.73	0.48
39	39	0.67	0.78	0.86	0.90	0.95	0.95	-0.02	-0.01	0.02	0.00	0.02	-0.01	0.00	-2.05	0.01	1.00	112	1.04	0.26
40	40	0.67	0.50	0.77	0.80	0.73	0.95	0.17	-0.13	0.06	0.00	-0.14	0.03	0.04	2.37	0.77	1.08	166	0.70	0.24
41	41	0.33	0.61	0.73	0.90	1.00	1.00	-0.20	-0.05	-0.01	0.08	0.12	0.07	0.00	2.75	-1.77	0.81	157	1.71	0.55
42	42	0.38	0.56	0.64	0.90	0.77	0.95	-0.06	-0.02	-0.03	0.14	-0.06	0.04	0.00	0.90	-0.73	0.92	179	1.19	0.43
43	43	0.19	0.56	0.55	0.60	0.59	0.90	-0.10	0.15	0.04	-0.01	-0.13	0.07	0.00	1.61	0.06	1.00	205	1.03	0.38
44	44	0.14	0.56	0.41	0.50	0.91	0.76	-0.13	0.17	-0.07	-0.09	0.21	-0.06	0.00	3.06	-0.40	0.96	206	1.15	0.43
45	45	0.67	0.17	0.23	0.25	0.23	0.14	0.57	0.02	0.03	-0.03	-0.16	-0.42	0.22	11.80	4.01	1.49	170	-0.93	-0.26
46	46	0.29	0.22	0.18	0.00	0.18	0.33	0.23	0.13	0.05	-0.19	-0.10	-0.11	0.09	6.47	1.54	1.19	138	0.05	0.00
47	47	0.19	0.72	0.68	0.85	0.91	0.90	-0.26	0.14	0.01	0.08	0.07	-0.01	0.00	2.72	-0.96	0.90	177	1.40	0.45
48	48	0.43	0.39	0.45	0.75	0.77	0.67	0.13	-0.03	-0.07	0.12	0.04	-0.18	0.05	2.63	0.97	1.10	203	0.59	0.26
49	49	0.62	0.72	0.64	0.65	0.91	0.76	0.15	0.13	-0.05	-0.12	0.06	-0.15	0.08	3.55	1.49	1.16	174	0.37	0.14
50	50	0.05	0.11	0.00	0.10	0.05	0.33	0.02	0.07	-0.06	0.01	-0.10	0.07	0.00	1.90	-0.16	0.97	84	0.96	0.24

RASCH 47

This is ONE-PARAMETER TEST ITEM ANALYSIS PROGRAM

(Rasch Model), Version 8 (2547)

to analyze M/C item.

It was originally written for DOS by

Dr. Suphat Sukamolson, Ph.D. *

(Educational Measurement and Evaluation),

and updated for Windows by

Mr. Komson Eiamjamrus, M.Sc.

LANGUAGE INSTITUTE

CHULALONGKORN UNIVERSITY, 2004

ALL RIGHTS RESERVED

[* Sukamolson, S. *Computerized Item Analyses and Grading*. BKK: Wittayaphat, 2542.]

PROCEDURE IS UCON**DIFFICULTY SCALE FACTOR 1.19 ABILITY SCALE FACTOR 1.15****NUMBER OF ITERATIONS = 2**

SEQUENCE NUMBER	ITEM NAME	ITEM DIFFICULTY	STANDARD ERROR	LAST DIFF CHANGE	PROX DIFF	FIRST CYCLE
1	1	0.261	0.139	0.001	0.279	0.260
2	2	1.441	0.159	0.003	1.510	1.438
3	3	2.384	0.205	0.004	2.520	2.381
4	4	0.853	0.145	0.002	0.895	0.850
5	5	-0.774	0.149	-0.002	-0.811	-0.772
6	6	-0.239	0.141	-0.001	-0.242	-0.238
7	7	-0.687	0.147	-0.002	-0.717	-0.685
8	8	1.155	0.151	0.003	1.210	1.153
9	9	0.492	0.140	0.001	0.520	0.491
10	10	-0.397	0.142	-0.001	-0.409	-0.396
11	11	-0.007	0.139	0.000	0.001	-0.007
12	12	0.937	0.146	0.003	0.982	0.934
13	13	-0.045	0.139	0.000	-0.039	-0.045
14	14	-0.519	0.144	-0.001	-0.538	-0.518
15	15	0.610	0.141	0.002	0.642	0.608
16	16	-0.045	0.139	0.000	-0.039	-0.045
17	17	0.979	0.147	0.003	1.026	0.976
18	18	-0.239	0.141	-0.001	-0.242	-0.238
19	19	-0.200	0.140	0.000	-0.201	-0.199
20	20	0.261	0.139	0.001	0.279	0.260
21	21	-0.278	0.141	-0.001	-0.284	-0.277
22	22	-0.478	0.144	-0.001	-0.495	-0.477
23	23	0.414	0.140	0.001	0.439	0.413
24	24	1.110	0.150	0.003	1.163	1.108
25	25	1.247	0.154	0.003	1.306	1.244
26	26	0.649	0.142	0.002	0.683	0.648
27	27	1.651	0.167	0.003	1.732	1.648
28	28	-0.397	0.142	-0.001	-0.409	-0.396
29	29	0.070	0.139	0.000	0.081	0.069
30	30	-0.084	0.139	0.000	-0.080	-0.083
31	31	0.492	0.140	0.001	0.520	0.491
32	32	-1.482	0.174	-0.003	-1.582	-1.478
33	33	-0.774	0.149	-0.002	-0.811	-0.772
34	34	-0.644	0.146	-0.002	-0.672	-0.643
35	35	0.108	0.139	0.000	0.120	0.108
36	36	-0.045	0.139	0.000	-0.039	-0.045
37	37	-0.161	0.140	0.000	-0.161	-0.160
38	38	-1.004	0.155	-0.002	-1.058	-1.001
39	39	-1.963	0.201	-0.004	-2.121	-1.958
40	40	-0.819	0.150	-0.002	-0.858	-0.816
41	41	-0.200	0.140	0.000	-0.201	-0.199
42	42	0.853	0.145	0.002	0.895	0.850

SEQUENCE NUMBER	ITEM NAME	ITEM DIFFICULTY	STANDARD ERROR	LAST DIFF CHANGE	PROX DIFF	FIRST CYCLE
43	43	-0.687	0.147	-0.002	-0.717	-0.685
44	44	0.811	0.144	0.002	0.852	0.809
45	45	0.414	0.140	0.001	0.439	0.413
46	46	0.299	0.139	0.001	0.319	0.298
47	47	0.222	0.139	0.001	0.240	0.222
48	48	-1.482	0.174	-0.003	-1.582	-1.478
49	49	1.597	0.165	0.003	1.675	1.594
50	50	-0.819	0.150	-0.002	-0.858	-0.816
51	51	2.761	0.233	0.004	2.935	2.757
52	52	0.531	0.141	0.002	0.560	0.529
53	53	-0.007	0.139	0.000	0.001	-0.007
54	54	-0.122	0.140	0.000	-0.120	-0.122
55	55	-1.152	0.160	-0.003	-1.219	-1.149
56	56	-0.397	0.142	-0.001	-0.409	-0.396
57	57	0.979	0.147	0.003	1.026	0.976
58	58	0.070	0.139	0.000	0.081	0.069
59	59	1.492	0.161	0.003	1.564	1.488
60	60	-0.007	0.139	0.000	0.001	-0.007
61	61	-0.084	0.139	0.000	-0.080	-0.083
62	62	-0.357	0.142	-0.001	-0.367	-0.356
63	63	0.222	0.139	0.001	0.240	0.222
64	64	0.337	0.139	0.001	0.359	0.336
65	65	0.649	0.142	0.002	0.683	0.648
66	66	-1.004	0.155	-0.002	-1.058	-1.001
67	67	0.299	0.139	0.001	0.319	0.298
68	68	-0.278	0.141	-0.001	-0.284	-0.277
69	69	0.261	0.139	0.001	0.279	0.260
70	70	0.610	0.141	0.002	0.642	0.608
71	71	0.811	0.144	0.002	0.852	0.809
72	72	-0.007	0.139	0.000	0.001	-0.007
73	73	0.570	0.141	0.002	0.601	0.569
74	74	0.070	0.139	0.000	0.081	0.069
75	75	-0.045	0.139	0.000	-0.039	-0.045
76	76	0.414	0.140	0.001	0.439	0.413
77	77	1.342	0.156	0.003	1.406	1.339
78	78	0.261	0.139	0.001	0.279	0.260
79	79	0.492	0.140	0.001	0.520	0.491
80	80	0.811	0.144	0.002	0.852	0.809
81	81	-0.602	0.146	-0.002	-0.627	-0.600
82	82	-1.101	0.159	-0.003	-1.164	-1.099
83	83	0.453	0.140	0.001	0.479	0.452
84	84	-0.317	0.141	-0.001	-0.325	-0.317
85	85	-1.739	0.187	-0.004	-1.869	-1.735
86	86	-1.671	0.183	-0.004	-1.792	-1.667
87	87	-0.239	0.141	-0.001	-0.242	-0.238
88	88	-1.482	0.174	-0.003	-1.582	-1.478
89	89	-1.963	0.201	-0.004	-2.121	-1.958

SEQUENCE NUMBER	ITEM NAME	ITEM DIFFICULTY	STANDARD ERROR	LAST DIFF CHANGE	PROX DIFF	FIRST CYCLE
90	90	1.342	0.156	0.003	1.406	1.339
91	91	-0.819	0.150	-0.002	-0.858	-0.816
92	92	-0.478	0.144	-0.001	-0.495	-0.477
93	93	-1.963	0.201	-0.004	-2.121	-1.958
94	94	-0.317	0.141	-0.001	-0.325	-0.317
95	95	0.414	0.140	0.001	0.439	0.413
96	96	-0.239	0.141	-0.001	-0.242	-0.238
97	97	-0.910	0.153	-0.002	-0.956	-0.907
98	98	-0.730	0.148	-0.002	-0.764	-0.728
99	99	-1.810	0.191	-0.004	-1.948	-1.806
100	100	-0.200	0.140	0.000	-0.201	-0.199

ROOT MEAN SQUARE = 0.002

100 ITEMS CALIBRATED ON 240 PERSONS

240 MEASURABLE PERSONS WITH MEAN ABILITY = 0.20 AND STD. DEV. = 0.80

COMPLETE SCORE EQUIVALENCE TABLE

RAW SCORE	COUNT	LOG ABILITY	STANDARD ERRORS	TEST CHARACTERISTIC CURVE
11	0	-2.36	0.33	
10	0	-2.47	0.35	
9	0	-2.60	0.36	
8	0	-2.73	0.38	
7	0	-2.89	0.40	
6	0	-3.06	0.43	
5	0	-3.26	0.47	
4	0	-3.50	0.52	
3	0	-3.81	0.60	
2	0	-4.23	0.72	
1	0	-4.93	1.01	

PERSON SEPARABILITY INDEX 0.92 (EQUIVALENT TO KR20)

100 ITEMS CALIBRATED ON 240 PERSONS

240 MEASURABLE PERSONS WITH MEAN ABILITY = 0.20 AND STD. DEV. = 0.60

MAP OF VARIABLE

PERSON STATS	COUNT	RAW SCORE	MEASURE MIDPOINT (S.E.)	ITEM COUNTS	TYPICAL ITEMS (BY NAME)															
	2	93	2.90 (0.41)																	
+3SD		91	2.70 (0.37)	1	51															
		90	2.50 (0.35)																	
	2	88	2.30 (0.32)	1	3															
	2	86	2.10 (0.30)																	
+2SD	4	83	1.90 (0.28)																	
	4	81	1.70 (0.27)	1	27															
	6	78	1.50 (0.26)	3	2 49 59															
	8	75	1.30 (0.25)	3	25 77 90															
+1SD	12	71	1.10 (0.24)	2	8 24															
	16	67	0.90 (0.23)	8	4 12 17 42 44 57 71 80															
	20	63	0.70 (0.22)	4	15 26 65 70															
	20	59	0.50 (0.22)	10	9 23 31 45 52 73 76 79 83 95															
MEAN	14	55	0.30 (0.22)	9	1 20 46 47 63 64 67 69 78															
	24	51	0.10 (0.22)	4	29 35 58 74															
	22	46	-0.10 (0.22)	15	11 13 16 19 30 36 37 41 53 54															
	24	42	-0.30 (0.22)	12	6 10 18 21 28 56 62 68 84 87															
	22	38	-0.50 (0.22)	3	14 22 92															
-1SD	8	34	-0.70 (0.23)	7	5 7 33 34 43 81 98															
	12	30	-0.90 (0.23)	4	40 50 91 97															
	12	27	-1.10 (0.24)	4	38 55 66 82															
	4	23	-1.30 (0.25)																	
-2SD	2	20	-1.50 (0.27)	3	32 48 88															
		18	-1.70 (0.28)	2	85 86															
		15	-1.90 (0.30)	4	39 89 93 99															
		13	-2.10 (0.31)																	

100 ITEMS CALIBRATED ON 240 PERSONS

240 MEASURABLE PERSONS WITH MEAN ABILITY = 0.20 AND STD. DEV. = 0.80

		ITEM CHARACTERISTIC CURVE						DEPARTURE FROM EXPECTED ICC						ITEM FIT STATISTICS						
SEQ	ITEM	1ST	2ND	3RD	4TH	5TH	6TH	1ST	2ND	3RD	4TH	5TH	6TH	* ERR	FIT	T-TESTS	WEIGHTED	DISC	POINT	
NUM	NAME	GROUP	GROUP	GROUP	GROUP	GROUP	GROUP	GROUP	GROUP	GROUP	GROUP	GROUP	GROUP	IMPAC	BETWN	TOTAL	MNSQ	DF	INDEX	BISER
1	1	0.11	0.21	0.70	0.52	0.45	0.76	-0.12	-0.12	0.29	0.11	-0.17	0.00	0.00	3.73	0.04	1.00	204	1.06	0.37
2	2	0.16	0.11	0.30	0.19	0.25	0.48	0.07	-0.03	0.12	-0.05	-0.08	-0.04	0.04	1.29	0.69	1.08	155	0.69	0.24
3	3	0.05	0.05	0.10	0.19	0.10	0.24	0.02	0.00	0.02	0.08	-0.06	-0.06	0.01	0.30	0.20	1.02	94	0.71	0.21
4	4	0.00	0.16	0.50	0.48	0.40	0.57	-0.14	-0.06	0.22	0.11	-0.07	-0.08	0.02	3.12	0.51	1.05	188	0.99	0.33
5	5	0.58	0.42	0.75	0.71	0.85	0.86	0.13	-0.16	0.09	-0.03	0.03	-0.04	0.03	1.33	0.54	1.05	177	0.82	0.26
6	6	0.16	0.21	0.60	0.71	0.85	0.95	-0.17	-0.24	0.07	0.08	0.13	0.11	0.00	3.35	-2.08	0.80	200	1.78	0.59
7	7	0.47	0.42	0.75	0.67	0.90	0.86	0.04	-0.14	0.11	-0.06	0.10	-0.03	0.00	1.26	0.03	1.00	182	1.00	0.33
8	8	0.05	0.32	0.25	0.29	0.25	0.62	-0.06	0.15	0.03	-0.21	-0.15	0.04	0.02	1.64	0.33	1.03	173	0.91	0.31
9	9	0.32	0.47	0.40	0.33	0.50	0.57	0.13	0.19	0.04	-0.12	-0.06	-0.15	0.10	2.87	2.02	1.21	200	0.27	0.12
10	10	0.16	0.47	0.70	0.71	0.80	0.86	-0.20	-0.02	0.13	0.05	0.04	0.00	0.00	1.54	-0.71	0.93	195	1.33	0.43
11	11	0.42	0.42	0.35	0.52	0.75	0.76	0.14	0.03	-0.13	-0.35	0.07	-0.05	0.04	1.15	0.84	1.08	204	0.76	0.26
12	12	0.16	0.42	0.40	0.33	0.35	0.38	0.03	0.22	0.14	-0.01	-0.10	-0.25	0.09	3.97	1.76	1.19	184	0.14	0.13
13	13	0.32	0.32	0.30	0.71	0.75	0.86	0.03	-0.09	-0.19	0.13	0.07	0.04	0.00	1.65	-0.54	0.94	203	1.26	0.43
14	14	0.32	0.58	0.65	0.52	0.85	0.95	-0.08	0.06	0.05	-0.17	0.07	0.08	0.00	1.66	-0.38	0.96	190	1.24	0.39
15	15	0.21	0.53	0.50	0.43	0.50	0.29	0.04	0.27	0.17	0.00	-0.03	-0.41	0.15	6.04	3.00	1.33	197	-0.17	0.00
16	16	0.63	0.37	0.45	0.48	0.65	0.71	0.35	-0.04	-0.04	-0.11	-0.03	-0.10	0.08	3.89	1.60	1.16	203	0.26	0.17
17	17	0.21	0.26	0.20	0.38	0.20	0.71	0.08	0.07	-0.05	0.04	-0.24	0.09	0.03	2.37	0.55	1.05	182	0.84	0.27
18	18	0.58	0.63	0.40	0.57	0.70	0.57	0.25	0.18	-0.13	0.03	-0.02	-0.27	0.11	5.17	2.12	1.22	200	-0.05	0.07
19	19	0.58	0.53	0.40	0.43	0.80	0.76	0.26	0.08	-0.12	-0.19	0.08	-0.07	0.07	3.62	1.34	1.14	200	0.45	0.20
20	20	0.21	0.58	0.60	0.29	0.70	0.52	-0.02	0.25	0.19	-0.23	0.08	-0.24	0.09	4.98	1.75	1.18	204	0.30	0.16
21	21	0.42	0.58	0.60	0.57	0.65	0.57	0.09	0.12	0.06	0.03	-0.08	-0.18	0.07	2.48	1.48	1.15	198	0.32	0.16
22	22	0.16	0.47	0.50	0.86	0.95	0.86	-0.22	-0.04	-0.09	0.17	0.18	-0.01	0.00	3.35	-1.74	0.83	191	1.59	0.54
23	23	0.42	0.32	0.30	0.38	0.45	0.81	0.22	0.02	-0.07	-0.09	-0.13	0.07	0.03	2.68	0.61	1.06	202	0.71	0.29
24	24	0.05	0.26	0.15	0.14	0.45	0.76	-0.06	0.09	-0.08	-0.17	0.04	0.17	0.00	2.44	-1.13	0.88	175	1.41	0.47
25	25	0.21	0.21	0.15	0.33	0.30	0.48	0.11	0.05	-0.06	0.05	-0.07	-0.08	0.04	1.35	0.78	1.08	167	0.57	0.23
26	26	0.21	0.26	0.25	0.43	0.50	0.71	0.04	0.01	-0.07	0.01	-0.02	0.02	0.00	-1.26	-0.01	1.00	196	0.98	0.36
27	27	0.16	0.16	0.15	0.10	0.15	0.57	0.09	0.05	0.00	-0.11	-0.14	0.11	0.00	2.22	-0.38	0.95	142	0.90	0.30
28	28	0.42	0.58	0.75	0.71	0.55	0.71	0.06	0.09	0.18	0.05	-0.21	-0.15	0.06	3.50	1.18	1.12	195	0.30	0.19
29	29	0.21	0.53	0.35	0.38	0.70	0.95	-0.05	0.15	-0.11	-0.18	0.04	0.15	0.00	2.75	-0.62	0.94	204	1.30	0.44
30	30	0.21	0.47	0.25	0.62	0.75	1.00	-0.08	0.06	-0.24	0.02	0.06	0.18	0.00	3.19	-1.34	0.87	203	1.58	0.52
31	31	0.26	0.21	0.10	0.52	0.60	0.86	0.07	-0.07	-0.26	0.07	0.04	0.13	0.00	2.84	-1.05	0.90	200	1.35	0.47
32	32	0.47	0.74	0.80	0.90	0.95	1.00	-0.15	0.00	0.00	0.05	0.05	0.05	0.00	1.01	-0.92	0.89	131	1.54	0.42
33	33	0.21	0.58	0.80	0.76	0.90	0.90	-0.24	-0.01	0.14	0.02	0.08	0.00	0.00	2.22	-0.99	0.90	177	1.44	0.44
34	34	0.26	0.32	0.90	0.76	0.85	0.90	-0.16	-0.24	0.27	0.04	0.05	0.02	0.00	3.70	-1.31	0.87	184	1.45	0.48
35	35	0.11	0.26	0.45	0.62	0.80	0.81	-0.15	-0.11	0.00	0.07	0.15	0.02	0.00	1.68	-1.36	0.87	205	1.48	0.52
36	36	0.37	0.32	0.60	0.57	0.70	0.71	0.08	-0.09	0.11	-0.01	0.02	-0.10	0.04	0.92	0.79	1.08	203	0.71	0.26
37	37	0.26	0.53	0.45	0.52	0.70	0.95	-0.05	0.09	-0.06	-0.09	-0.01	0.12	0.00	1.10	-0.34	0.96	201	1.24	0.40
38	38	0.58	0.58	0.55	0.81	0.90	1.00	0.07	-0.06	-0.16	0.02	0.05	0.08	0.00	1.64	-0.35	0.96	163	1.25	0.39
39	39	0.79	0.79	0.85	0.90	0.90	1.00	0.06	-0.03	-0.01	0.00	-0.04	0.03	0.00	-0.36	-0.10	0.98	98	0.95	0.24
40	40	0.42	0.58	0.50	0.76	1.00	0.95	-0.04	-0.02	-0.17	0.01	0.18	0.05	0.00	2.36	-0.73	0.92	174	1.43	0.44
41	41	0.21	0.37	0.45	0.67	0.85	0.90	-0.11	-0.07	-0.07	0.04	0.13	0.07	0.00	1.25	-1.33	0.87	200	1.51	0.51
42	42	0.26	0.26	0.10	0.24	0.45	0.81	0.12	0.05	-0.18	-0.13	-0.02	0.16	0.00	2.94	-0.57	0.94	188	1.12	0.41
43	43	0.32	0.68	0.55	0.90	0.75	0.86	-0.12	0.12	-0.09	0.18	-0.05	-0.03	0.00	2.11	-0.17	0.98	182	1.04	0.34
44	44	0.32	0.21	0.25	0.33	0.35	0.71	0.17	-0.01	-0.04	-0.04	-0.13	0.06	0.03	1.95	0.58	1.06	190	0.77	0.27
45	45	0.26	0.21	0.15	0.81	0.45	0.76	0.06	-0.09	-0.22	0.34	-0.13	0.02	0.00	4.36	-0.01	1.00	202	1.08	0.37
46	46	0.26	0.37	0.40	0.62	0.50	0.67	0.04	0.04	0.00	0.12	-0.11	-0.09	0.04	0.77	0.77	1.07	204	0.66	0.28
47	47	0.47	0.47	0.45	0.57	0.50	0.48	0.24	0.13	0.03	0.05	-0.12	-0.30	0.13	4.95	2.62	1.28	204	-0.15	0.04
48	48	0.47	0.58	0.90	1.00	0.85	0.95	-0.15	-0.06	0.10	0.14	-0.05	0.00	0.00	2.34	-0.39	0.95	131	1.33	0.32
49	49	0.05	0.05	0.05	0.24	0.35	0.57	-0.02	-0.06	-0.10	0.02	0.05	0.09	0.00	0.86	-0.78	0.91	145	1.41	0.45
50	50	0.42	0.42	0.75	0.86	0.85	0.90	-0.04	-0.17	0.08	0.10	0.03	0.00	0.00	1.19	-0.37	0.96	174	1.24	0.38

ภาคผนวก ก. ผลการวิเคราะห์ข้อทดสอบรายชื่อของแบบทดสอบทั้ง 3 ชุด

Difficulty, Discrimination and Test Item Validity Indexes of TEPOIC Form 1.0

Item no.	Objective	p	d	r_{pb}	Comments
1		0.440	0.065	0.079	Low discrimination
2		0.440	0.548	0.500	
3		0.207	0.161	0.216	
4		0.836	0.226	0.313	
5		0.871	0.387	0.713	Too easy
6		0.483	0.387	0.319	
7		0.802	0.194	0.327	
8		0.310	0.323	0.250	
9		0.664	0.323	0.358	
10		0.888	0.290	0.509	Too easy
11		0.362	0.484	0.435	
12		0.517	0.323	0.367	
13		0.517	0.258	0.222	
14		0.853	0.290	0.493	
15		0.534	0.323	0.354	
16		0.543	0.484	0.451	
17		0.414	0.516	0.429	
18		0.836	0.419	0.608	
19		0.491	0.290	0.231	
20		0.414	0.387	0.350	
21		0.310	0.032	0.096	Low discrimination
22		0.621	0.323	0.311	
23		0.828	0.129	0.244	
24		0.690	0.516	0.601	
25		0.500	0.194	0.246	
26		0.440	0.290	0.267	
27		0.690	0.416	0.527	
28		0.457	0.194	0.213	
29		0.888	0.258	0.422	Too easy
30		0.750	0.258	0.284	
31		0.810	0.355	0.410	
32		0.733	0.484	0.616	
33		0.681	0.323	0.328	
34		0.819	0.516	0.726	
35		0.733	0.645	0.673	
36		0.647	0.613	0.571	
37		0.595	0.161	0.248	
38		0.310	0.355	0.473	
39		0.647	0.516	0.515	
40		0.690	0.258	0.309	
41		0.578	0.355	0.305	
42		0.310	0.355	0.405	
43		0.595	0.548	0.516	
44		0.491	0.452	0.441	
45		0.733	0.258	0.279	
46		0.457	0.452	0.447	
47		0.336	0.065	0.047	Low discrimination
48		0.474	0.387	0.366	

49		0.414	0.258	0.204	
50		0.931	0.194	0.605	Too easy
51		0.500	0.484	0.474	
52		0.569	0.355	0.313	
53		0.638	0.194	0.220	
54		0.328	0.258	0.355	
55		0.302	0.516	0.476	
56		0.690	0.065	0.042	Low discrimination
57		0.474	0.452	0.527	
58		0.586	0.581	0.593	
59		0.698	0.516	0.567	
60		0.371	0.290	0.343	
61		0.484	0.260	0.330	
62		0.828	0.194	0.315	
63		0.647	0.516	0.588	
64		0.328	0.226	0.342	
65		0.664	0.416	0.410	
66		0.388	0.097	0.172	Low discrimination
67		0.276	0.258	0.305	
68		0.336	0.323	0.355	
69		0.595	0.258	0.337	
70		0.414	0.226	0.314	
71		0.491	0.677	0.701	
72		0.233	0.226	0.259	
73		0.612	0.581	0.554	
74		0.233	0.129	0.115	
75		0.517	0.677	0.572	
76		0.172	0.258	0.422	
77		0.733	0.484	0.463	
78		0.233	0.000	0.017	
79		0.441	0.323	0.313	
80		0.371	0.290	0.394	
81		0.940	0.000	0.129	Low discrimination
82		0.914	0.161	0.353	Too easy
83		0.586	0.484	0.516	
84		0.534	0.161	0.192	
85		0.681	0.355	0.390	
86		0.672	0.516	0.578	
87		0.422	0.161	0.161	
88		0.948	0.161	0.544	Too easy
89		0.759	0.290	0.402	
90		0.922	0.129	0.333	Too easy
91		0.491	0.258	0.272	
92		0.914	0.097	0.187	Too easy
93		0.690	0.387	0.413	
94		0.543	0.452	0.428	
95		0.836	0.419	0.626	
96		0.466	0.065	0.109	Low discrimination
97		0.603	0.710	0.647	
98		0.672	0.613	0.545	
99		0.569	0.645	0.599	
100		0.819	0.355	0.468	N = 15

Difficulty, Discrimination and Test Item Validity Indexes of TEPOIC Form 2.0

Item no.	Objective	p	d	r_{pb}	Comments
1		0.653	0.576	0.578	
2		0.218	0.242	0.304	
3		0.629	0.606	0.570	
4		0.397	0.576	0.538	
5		0.040	0.000	0.048	Poor item.
6		0.516	0.242	0.248	
7		0.710	0.303	0.383	
8		0.669	0.667	0.636	
9		0.653	0.667	0.638	
10		0.653	0.606	0.607	
11		0.403	-0.121	-0.035	Poor item.
12		0.371	0.545	0.542	
13		0.677	0.455	0.483	
14		0.266	0.000	0.057	Poor item.
15		0.492	0.182	0.222	
16		0.492	0.636	0.567	
17		0.444	0.303	0.272	
18		0.444	0.212	0.262	
19		0.492	0.455	0.452	
20		0.395	0.424	0.435	
21		0.685	0.667	0.710	
22		0.258	0.182	0.180	Low discrimination
23		0.411	0.576	0.570	
24		0.637	0.636	0.630	
25		0.694	0.303	0.313	
26		0.823	0.364	0.599	
27		0.831	0.303	0.357	
28		0.331	0.636	0.536	
29		0.540	0.727	0.704	
30		0.613	0.606	0.623	
31		0.371	0.485	0.479	
32		0.298	0.303	0.420	
33		0.758	0.606	0.666	
34		0.515	0.667	0.599	
35		0.629	0.515	0.505	
36		0.718	0.313	0.336	
37		0.774	0.424	0.572	
38		0.839	0.455	0.663	
39		0.855	0.242	0.402	
40		0.742	0.242	0.312	
41		0.766	0.576	0.721	
42		0.702	0.455	0.543	
43		0.565	0.424	0.472	
44		0.548	0.485	0.534	
45		0.282	-0.303	-0.328	Poor item
46		0.202	-0.030	-0.000	Low discrimination
47		0.710	0.515	0.581	
48		0.581	0.303	0.328	

49		0.718	0.182	0.182	Low discrimination
50		0.105	0.182	0.364	Too difficult
51		0.452	0.424	0.455	
52		0.702	0.515	0.555	
53		0.637	0.485	0.518	
54		0.242	0.545	0.524	
55		0.774	0.152	0.217	
56		0.629	0.394	0.411	
57		0.677	0.515	0.535	
58		0.153	0.182	0.305	
59		0.589	0.576	0.568	
60		0.726	0.333	0.323	
61		0.234	0.303	0.409	
62		0.532	0.848	0.841	
63		0.573	0.697	0.628	
64		0.242	0.455	0.577	
65		0.419	0.727	0.694	
66		0.589	0.667	0.669	
67		0.460	0.152	0.121	
68		0.355	0.303	0.329	
69		0.226	0.152	0.218	
70		0.540	0.545	0.511	
71		0.569	0.636	0.676	
72		0.194	0.273	0.380	
73		0.218	0.394	0.536	
74		0.444	0.333	0.383	
75		0.427	0.515	0.525	
76		0.516	0.697	0.714	
77		0.315	0.242	0.332	
78		0.250	0.515	0.615	
79		0.274	0.212	0.331	
80		0.250	-0.121	-0.088	Low discrimination
81		0.782	0.273	0.364	
82		0.919	0.242	0.610	Too easy
83		0.718	0.576	0.591	
84		0.839	0.364	0.607	
85		0.919	0.182	0.508	Too easy
86		0.887	0.242	0.454	Too easy
87		0.806	0.424	0.586	
88		0.815	0.545	0.746	
89		0.782	0.606	0.691	
90		0.565	0.212	0.160	
91		0.556	0.303	0.317	
92		0.790	0.485	0.658	
93		0.742	0.606	0.690	
94		0.258	0.242	0.350	
95		0.444	0.788	0.737	
96		0.710	0.273	0.295	
97		0.556	0.697	0.649	
98		0.855	0.273	0.494	
99		0.532	0.364	0.390	
100		0.847	0.242	0.496	N = 12

Difficulty, Discrimination and Test Item Validity Indexes of TEPOIC Form 3.0

Item no.	Objectives	p	d	r_{pb}	Comments
1		0.483	0.563	0.463	
2		0.250	0.250	0.305	
3		0.125	0.125	0.307	Too difficult
4		0.358	0.375	0.413	
5		0.700	0.344	0.327	
6		0.592	0.719	0.730	
7		0.683	0.406	0.415	
8		0.300	0.344	0.392	
9		0.433	0.188	0.144	Low discrimination
10		0.625	0.563	0.538	
11		0.542	0.313	0.327	
12		0.342	0.125	0.162	Low discrimination
13		0.650	0.469	0.530	
14		0.650	0.531	0.488	
15		0.408	0.000	0.001	Low discrimination
16		0.550	0.188	0.207	
17		0.333	0.344	0.339	
18		0.592	0.031	0.091	Low discrimination
19		0.583	0.250	0.243	
20		0.483	0.188	0.197	Low discrimination
21		0.600	0.188	0.199	Low discrimination
22		0.642	0.625	0.683	
23		0.450	0.375	0.356	
24		0.308	0.563	0.597	
25		0.283	0.250	0.295	
26		0.400	0.469	0.449	
27		0.217	0.281	0.398	
28		0.625	0.250	0.236	
29		0.525	0.625	0.547	
30		0.558	0.656	0.644	
31		0.433	0.531	0.589	
32		0.817	0.406	0.586	
33		0.700	0.500	0.561	
34		0.675	0.625	0.614	
35		0.517	0.625	0.649	
36		0.550	0.406	0.328	
37		0.575	0.438	0.501	
38		0.742	0.406	0.506	
39		0.825	0.219	0.374	
40		0.708	0.469	0.563	
41		0.583	0.656	0.633	
42		0.358	0.438	0.507	
43		0.683	0.344	0.430	
44		0.367	0.313	0.343	
45		0.450	0.375	0.459	
46		0.475	0.281	0.349	
47		0.492	0.000	0.045	Low discrimination
48		0.817	0.375	0.450	

49		0.225	0.406	0.593	
50		0.708	0.500	0.487	
51		0.092	-0.156	-0.270	Too difficult
52		0.425	0.688	0.645	
53		0.542	0.531	0.555	
54		0.567	0.469	0.468	
55		0.767	0.375	0.479	
56		0.625	0.500	0.494	
57		0.333	0.469	0.515	
58		0.525	0.375	0.382	
59		0.242	0.063	0.176	Low discrimination
60		0.542	0.563	0.531	
61		0.558	0.625	0.672	
62		0.617	0.531	0.574	
63		0.492	0.719	0.688	
64		0.467	0.531	0.577	
65		0.400	0.156	0.189	Low discrimination
66		0.742	0.688	0.701	
67		0.475	0.625	0.649	
68		0.600	0.406	0.403	
69		0.483	0.594	0.634	
70		0.408	0.344	0.372	
71		0.367	0.438	0.420	
72		0.542	0.688	0.640	
73		0.417	0.688	0.635	
74		0.525	0.469	0.484	
75		0.550	0.438	0.448	
76		0.450	0.344	0.309	
77		0.267	0.125	0.172	Low discrimination
78		0.483	0.250	0.241	
79		0.433	0.625	0.569	
80		0.367	0.313	0.347	
81		0.667	0.500	0.564	
82		0.758	0.594	0.704	
83		0.442	0.031	0.082	Low discrimination
84		0.608	0.719	0.663	
85		0.850	0.406	0.686	
86		0.842	0.406	0.571	
87		0.592	0.438	0.468	
88		0.817	0.563	0.772	
89		0.855	0.438	0.745	
90		0.267	0.188	0.170	Low discrimination
91		0.708	0.438	0.428	
92		0.642	0.219	0.286	
93		0.875	0.188	0.405	Too easy
94		0.608	0.188	0.261	
95		0.450	0.594	0.588	
96		0.592	0.281	0.286	
97		0.725	0.656	0.787	
98		0.692	0.313	0.305	
99		0.858	0.250	0.491	
100		0.583	0.469	0.533	N = 15

ภาคผนวก ง. ตารางเปรียบเทียบผลการสอบแบบทดสอบทั้ง 3 ชุดกับ TOEIC

(SD-SER-4.6-481-R-0)

ตารางเทียบเคียงคะแนนมาตรฐานต่างประเทศ TOEIC

FORM 1				FORM 2				FORM 3			
CULI-TEST	TOEIC	CULI-TEST	TOEIC	CULI-TEST	TOEIC	CULI-TEST	TOEIC	CULI-TEST	TOEIC	CULI-TEST	TOEIC
1	-77	51	523	1	-4	51	535	1	27	51	557
2	-65	52	535	2	7	52	546	2	38	52	568
3	-53	53	547	3	18	53	556	3	48	53	579
4	-41	54	559	4	29	54	567	4	59	54	589
5	-29	55	571	5	39	55	578	5	69	55	600
6	-17	56	583	6	50	56	589	6	80	56	611
7	-5	57	595	7	61	57	599	7	91	57	621
8	7	58	607	8	72	58	610	8	101	58	632
9	19	59	619	9	82	59	621	9	112	59	642
10	31	60	631	10	93	60	632	10	122	60	653
11	43	61	643	11	104	61	643	11	133	61	664
12	55	62	655	12	115	62	653	12	144	62	674
13	67	63	667	13	126	63	664	13	154	63	685
14	79	64	679	14	136	64	675	14	165	64	695
15	91	65	691	15	147	65	686	15	176	65	706
16	103	66	703	16	158	66	696	16	186	66	717
17	115	67	715	17	169	67	707	17	197	67	727
18	127	68	727	18	179	68	718	18	207	68	738
19	139	69	739	19	179	69	729	19	218	69	748
20	151	70	751	20	201	70	739	20	229	70	759
21	163	71	763	21	212	71	750	21	239	71	770
22	175	72	775	22	222	72	761	22	250	72	780
23	187	73	787	23	233	73	772	23	260	73	791
24	199	74	799	24	244	74	783	24	271	74	802
25	211	75	811	25	255	75	793	25	282	75	812
26	223	76	823	26	266	76	804	26	292	76	823
27	235	77	835	27	276	77	815	27	303	77	833
28	247	78	847	28	287	78	826	28	313	78	844
29	259	79	859	29	298	79	836	29	324	79	855
30	271	80	871	30	309	80	847	30	335	80	865
31	283	81	883	31	319	81	858	31	345	81	876
32	295	82	895	32	330	82	869	32	356	82	886
33	307	83	907	33	341	83	879	33	367	83	897
34	319	84	919	34	352	84	890	34	377	84	908
35	331	85	931	35	362	85	901	35	388	85	918

36	343	86	943	36	373	86	912	36	398	86	929
37	355	87	955	37	384	87	923	37	409	87	939
38	367	88	967	38	395	88	933	38	420	88	950
39	379	89	979	39	406	89	944	39	430	89	961
40	391	90	991	40	416	90	955	40	441	90	971
41	403	91	1003	41	427	91	966	41	451	91	982
42	415	92	1015	42	438	92	976	42	462	92	993
43	427	93	1027	43	449	93	987	43	473	93	1003
44	439	94	1039	44	459	94	998	44	483	94	1014
45	451	95	1051	45	470	95	1009	45	494	95	1024
46	463	96	1063	46	481	96	1019	46	504	96	1035
47	475	97	1075	47	492	97	1030	47	515	97	1046
48	487	98	1087	48	503	98	1041	48	526	98	1056
49	499	99	1099	49	513	99	1052	49	536	99	1067
50	511	100	1111	50	524	100	1063	50	547	100	1077

หมายเหตุ ค่าความคลาดเคลื่อน FORM 1 = + 45

หมายเหตุ ค่าความคลาดเคลื่อน FORM 2 = + 68

หมายเหตุ ค่าความคลาดเคลื่อน FORM 3 = + 86

ภาคผนวก จ.

สำเนาปกของแบบทดสอบ CULI-TEPOIC ชุดที่ 1, 2 และ 3

ต่อไปนี้เป็นสำเนาปกของแบบทดสอบ CULI-TEPOIC ชุดที่ 1, 2 และ 3 ที่สร้างและพัฒนามาขึ้น ส่วนรายละเอียดของแบบทดสอบแต่ละชุดไม่สามารถเปิดเผยได้เนื่องจากเป็นแบบทดสอบที่จะนำไปใช้จริงในโอกาสต่อไป

L - FORM I

CHULALONGKORN UNIVERSITY
LANGUAGE INSTITUTE

*TEST OF ENGLISH FOR OCCUPATION AND INTERNATIONAL
COMMUNICATION*

LISTENING

TIME: 40 MINUTES
NUMBER OF PAGES: 12 PAGES
NUMBER OF QUESTIONS: 50 QUESTIONS

<u>PART</u>	¹⁹⁶ <u>QUESTION NUMBERS</u>
I	1 - 10
II	11 - 20
III	21 - 30
IV	31 - 40
V	41 - 50

ห้ามนำแบบทดสอบนี้ไปเผยแพร่โดยไม่ได้รับอนุญาตของสถาบัน

W - R - FORM I

CHULALONGKORN UNIVERSITY
LANGUAGE INSTITUTE

*TEST OF ENGLISH FOR OCCUPATION AND INTERNATIONAL
COMMUNICATION*

WRITING AND READING

TIME: 60 MINUTES
NUMBER OF PAGES: 14 PAGES
NUMBER OF QUESTIONS: 50 QUESTIONS

<u>PART</u>	<u>QUESTION NUMBERS</u>
VI	51 - 65
VII	66 - 80
VIII	81 - 100



CHULALONGKORN UNIVERSITY
LANGUAGE INSTITUTE

*TEST OF ENGLISH FOR PROFESSIONAL AND INTERNATIONAL
COMMUNICATION*

LISTENING

TIME: 40 MINUTES
NUMBER OF PAGES: 12 PAGES
NUMBER OF QUESTIONS: 50 QUESTIONS

<u>PART</u>	<u>QUESTION NUMBERS</u>
I	1 - 10
II	11 - 20
III	21 - 30
IV	31 - 40
V	41 - 50

ห้ามนำแบบทดสอบ และเอกสารใด ๆ ออกนอกห้องสอบ

W - R - FORM II

CHULALONGKORN UNIVERSITY
LANGUAGE INSTITUTE

*TEST OF ENGLISH FOR OCCUPATION AND INTERNATIONAL
COMMUNICATION*

WRITING AND READING

TIME: 60 MINUTES
NUMBER OF PAGES: 17 PAGES
NUMBER OF QUESTIONS: 50 QUESTIONS

<u>PART</u>	<u>QUESTION NUMBERS</u>
VI	51 - 65
VII	66 - 80
VIII	81 - 100



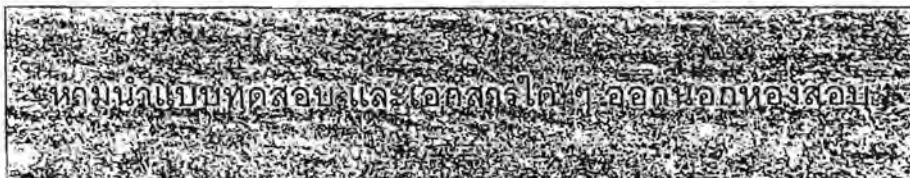
CHULALONGKORN UNIVERSITY
LANGUAGE INSTITUTE

*TEST OF ENGLISH FOR OCCUPATION AND INTERNATIONAL
COMMUNICATION*

LISTENING

TIME: 40 MINUTES
NUMBER OF PAGES: 12 PAGES
NUMBER OF QUESTIONS: 50 QUESTIONS

<u>PART</u>	<u>QUESTION NUMBERS</u>
I	1 - 10
II	11 - 20
III	21 - 30
IV	31 - 40
V	41 - 50



W - R - FORM III

CHULALONGKORN UNIVERSITY
LANGUAGE INSTITUTE

*TEST OF ENGLISH FOR OCCUPATION AND INTERNATIONAL
COMMUNICATION*

WRITING AND READING

TIME: 60 MINUTES
NUMBER OF PAGES: 14 PAGES
NUMBER OF QUESTIONS: 50 QUESTIONS

<u>PART</u>	<u>QUESTION NUMBERS</u>
VI	51 - 65
VII	66 - 80
VIII	81 - 100

ห้ามนำแบบทดสอบ และเอกสารใด ๆ ออกนอกห้องสอบ

ภาคผนวก ฉ.

ความสำคัญของภาษาอังกฤษในยุคของเทคโนโลยีเพื่อการสื่อสาร

ต่อไปนี้เป็นสำเนาเอกสารแสดงความสำคัญของภาษาอังกฤษในยุคของเทคโนโลยีเพื่อการสื่อสาร

Internet World Stats

Usage and Population Statistics



Index | World Stats | Africa Stats | America Stats | Asia Stats | Europe Stats | M. East Stats | Oceania Stats | Site Links

INTERNET USERS BY LANGUAGE

Mobile Internet Access

150 Countries Internet Access
Keep using your Email anywhere

Internet Statistics

Find the latest and greatest free
Internet Statistics info.

Internet Surveys

Online directory: Internet Surveys
& other marketing resources.

Sign Language

Over 900 sign language de-
T-shirts & more! Easy retur

Ads by Goc

Top Ten Languages Used in the Web (Number of Internet Users by Language)

TOP TEN LANGUAGES IN THE INTERNET	% of all Internet Users	Internet Users by Language	Internet Penetration by Language	Internet Growth for Language (2000 - 2005)	World Populat 2006 Estim for the Lan
English	30.0 %	312,924,679	27.8 %	128.1 %	1,125,664,3
Chinese	13.8 %	144,301,513	10.8 %	346.7 %	1,340,767,1
Japanese	8.3 %	86,300,000	67.2 %	83.3 %	128,389,1
Spanish	7.5 %	78,166,075	17.9 %	216.7 %	429,293,3
German	5.6 %	58,214,778	60.7 %	110.9 %	95,982,1
French	4.4 %	45,807,499	12.0 %	275.5 %	381,193,3
Korean	3.2 %	33,900,000	45.8 %	78.0 %	73,945,1
Portuguese	3.1 %	32,372,000	14.0 %	327.3 %	230,846,3
Italian	2.8 %	28,870,000	48.8 %	118.7 %	59,115,3
Russian	2.3 %	23,700,000	16.5 %	664.5 %	143,682,3
TOP TEN LANGUAGES	81.0 %	844,556,544	21.0 %	160.7 %	4,008,879,3
Rest of World Languages	19.0 %	198,548,342	8.0 %	436.0 %	2,490,817,3
WORLD TOTAL	100.0 %	1,043,104,886	16.0 %	189.0 %	6,499,697,3

(*) NOTES: (1) Internet Top Ten Languages Usage Stats were updated on June 30, 2006. (2) Internet Penetration is the ratio between the sum of Internet users speaking a language and the total population estimate that speaks that specific language. (3) The most recent

Internet usage information comes from data published by [Nielsen//NetRatings](#), [International Telecommunications Union](#), [Computer](#)

[Industry Almanac](#), and other reliable sources. (4) World population information comes from the [world gazetteer](#) web site. (5) For definitions

and navigation help, see the [Site Surfing Guide](#). (6) Stats may be cited, stating the source and establishing an active link back to [Internet World Stats](#). ©Copyright 2006, Miniwatts Marketing Group. All rights reserved.

Examples for data interpretation:

- There are 78,166,075 Spanish speaking people using the Internet, this represents 7.5 % of all the Internet users.
- Out of the estimated 429,293,261 world population that speaks Spanish, 17.9 % use the Internet.
- The number of Spanish Speaking Internet Users has grown 216.7 % in the last five years (2000-2005).

Tallying the number of speakers of the world's languages is an increasingly complex task, particularly with the push in many countries to teach English in their public

Vistawide World Languages & Cultures

25 Language Schools

Languages Courses around the World
Spanish - French - German - Italian

Chinese courses in China

Looking to speak Chinese In Beijing or Shanghai? Visit our site!

Study English in USA

Intensive English, ESL, homestay
Toefl, F1 Visa, Small Class

Costa Rica Spanish Scho

Quality immersion study, all in 3 diverse sites -town-forest-b

Ads by Goooo

Languages

Language statistics
Languages in the U.S.
Why learn a language?
10 Language learning tips
15 Language resources
Language gift ideas
Free language learning software
Free foreign language fonts
Typing special characters
More ...

Ads by Google

Translate Languages
English Second Language
Papua New Guinea Airfares
Translate English to Italian
Language Translations

Books

Search for more books at amazon:



French

German

Italian

Japanese

Spanish

Google

Language statistics & facts

Language statistics & facts | [Top 30 languages](#) | [20 countries with the most languages](#)

Language families: [By speakers](#) - [By languages](#) | [Languages in the United States](#)

Number of living languages: 6912

Number of those languages that are nearly extinct: 516

Language with the greatest number of native speakers: **Mandarin Chinese** [See [Top 30 languages](#)]

Language spoken by the greatest number of non-native speakers: **English** (250 million to 350 million non-native speakers)

Country with the most languages spoken: **Papua New Guinea** has 820 living languages. [See [Top 20 countries](#)]

How long have languages existed: **Since about 100,000 BC**

First language ever written: **Sumerian** or **Egyptian** (about 3200 BC)

Oldest written language still in existence: **Chinese** or **Greek** (about 1500 BC)

Language with the most words: **English**, approx. 250,000 distinct words

Language with the fewest words: **Taki Taki** (also called **Sranan**), 340 words. Taki Taki is an English-based Creole spoken by 120,000 in the South American country of Suriname.

Language with the largest alphabet (including diacritics): **Kabardian** (58 letters). Kabardian belongs to the North Caucasian language family. The approx. 1 million Kabardian-speakers live primarily in Russia.

Language with the largest alphabet (graphically independent letters): **Georgian** (41 letters). This language of the Kartvelian family is spoken by over 4 million people primarily living in Georgia.

Language with the shortest alphabet: **Rotokas** (12 letters). Approx. 4300 people speak this East Papuan language. They live primarily in the Bougainville Province of Papua New Guinea.

Web vistawide.com**Forum****Career resources****Study abroad****Grants & scholarships**

The language with the fewest sounds (phonemes): **Rotokas** (11 phonemes)

The language with the most sounds (phonemes): **!Xóǀ** (112 phonemes). Approx. 4200 speak !Xóǀ, the vast majority of whom live in the African country of Botswana.

Language with the fewest consonant sounds: **Rotokas** (6 consonants)

Language with the most consonant sounds: **Ubyx** (81 consonants). This language of the North Caucasian Language family, once spoken in the Haci Osman village near Istanbul, has been extinct since 1992. Among living languages, **!Xóǀ** has the most consonants (77).

Language with the fewest vowel sounds: **Ubyx** (2 vowels). The related language **Abkhaz** also has 2 vowels in some dialects. There are approximately 106,000 Abkhaz speakers living primarily in Georgia.

Language with the most vowel sounds: **!Xóǀ** (31 vowels)

The most widely published language: **English**

Language with the fewest irregular verbs: **Esperanto** (none)

Language which has won the most Oscars: **Italian** (12 Academy Awards for Best Foreign Film)

The most translated document: **Universal Declaration Of Human Rights**, written by the United Nations in 1948, has been translated into 321 languages and dialects.

The most common consonant sounds in the world's languages: /p/, /t/, /k/, /m/, /n/

Longest word in the English language:
pneumonoultramicroscopicsilicovolcanoconiosis (45 letters)

30 Languages Available - Tap your innate language-learning ability. The Fastest Way to Learn rosetastone.com

Ads by Google

Sources Consulted: *Ethnologue: Languages of the World*, 15th ed. (2005). - *Oxford English Dictionary* - *Wikipedia.org* - *Guinness Book of World Records* (2005)

Language statistics | [Top 30 languages](#) | [20 countries with the most languages](#)

Language families: [By speakers](#) - [By languages](#) | [Languages in the United States](#)

Vistawide World Languages & Cultures

25 Language Schools

Languages Courses around the World Spanish - French - German - Italian

Learn Chinese in Beijing

Learn Chinese Language in Beijing Affordable & flexible programs

ESL Study Guide

Looking for speaking practice? get actual speaking practice

Learn Portuguese in Bah

Language courses in Salvador Brazil, starting at only \$170/w

Ads by Goood

Languages

- [Top 30 languages](#)
- [Languages in the U.S.](#)
- [Why learn a language?](#)
- [10 Language learning tips](#)
- [15 Language resources](#)
- [Language gift ideas](#)
- [Free language learning software](#)
- [Free foreign language fonts](#)
- [Typing special characters](#)
- [More...](#)

Ads by Google

- [Study Languages](#)
- [Learn a Language](#)
- [Learn Arabic](#)
- [English Language](#)
- [Italian Language Resources](#)

Books

Search for more books at amazon:



[French](#)

[German](#)

[Italian](#)

[Japanese](#)

[Spanish](#)

Google

Top 30 languages of the world

[Language statistics & facts](#) | [Top 30 languages](#) | [20 countries with the most languages](#)

[Language families: By speakers - By languages](#) | [Languages in the United States](#)

There are approximately 6900 languages currently spoken around the world, the majority of which have only a small number of speakers. About 4 billion of the earth's 6.5 billion people, or over 60% of the earth's population, speak one of the following 30 languages as their native tongue.

Languages are ordered in the table below by numbers of native speakers. Numbers of second-language speakers are given where known. Second language speakers are those whose native language is typically a minority language of the country in which they live and who learn the second language because it is the official national language of the country where they reside.

Top 30 Languages by Number of Native Speakers

Data source: *Ethnologue: Languages of the World*, 15th ed. (2005) & Wikipedia.org

Language	approximate # of speakers	Where is it spoken as an official language?
1. Mandarin Chinese	NATIVE: 873 million 2nd: 178 million TOTAL: 1,051 billion	OFFICIAL: People's Republic of China, Republic of China, Singapore
2. Hindi	NATIVE: 370 million 2nd: 120 million TOTAL: 490 million	OFFICIAL: India, Fiji
3. Spanish	NATIVE: 350 million 2nd: 70 million TOTAL: 420 million	OFFICIAL: Argentina, Bolivia, Chile, Colombia, Costa Rica, Cuba, Dominican Republic, Ecuador, El Salvador, Equatorial Guinea, Guatemala, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, Spain, United States (New Mexico, Puerto Rico, Uruguay, Venezuela
4. English	NATIVE: 340 million TOTAL: 510 million	OFFICIAL: Antigua and Barbuda, Australia, The Bahamas, Bangladesh, Barbados, Belize, Botswana, Brunei, Cameroon, Canada, Dominica, Ethiopia,

Web vistawide.com

Search

Eritrea, Fiji, The Gambia, Ghana, Grenada, Guyana, Hong Kong (People's Republic of China), India, Ireland, Jamaica, Kenya, Kiribati, Lesotho, Liberia, Malawi, Maldives, Malta, Marshall Islands, Maritius, Micronesia, Namibia, Nauru, New Zealand, Nigeria, Pakistan, Palau, Papua New Guinea, Philippines, Rwanda, Saint Kitts and Nevs, Saint Lucia, Saint Vincent and the Grenadines, Samoa, Seychelles, Sierra Leone, Singapore, Solomon Islands, Somolia, South Africa, Sri Lanka, Swaziland, Tanzania, Tonga, Trinidad and Tobago, Tuvalu, Uganda, United Kingdom, United States, Vanuatu, Zambia, Zimbabwe.

[Forum](#)

[Career resources](#)

[Study abroad](#)

[Grants & scholarships](#)



Learn a New LANGUAGE.

For a TRIP

For Your CAREER

OR

Just for FUN.

GET FREE 2-DAY SHIPPING*.
(*US only)

5. Arabic	NATIVE: 206 million 2nd: 24 million TOTAL: 230 million [World Almanac est. total 255 million]	OFFICIAL: Modern Standard Arabic: Algeria, Bahrain, Chad, Comoros, Djibouti, Egypt, Eritrea, Iraq, Israel, Jordan, Kuwait, Lebanon, Libya, Morocco, Niger, Oman, Palestinian Territories, Qatar, Saudi Arabia, Somalia, Sudan, Syria, Tunisia, United Arab Emirates, Western Sahara, Yemen. Hasaniya Arabic: Mauritania, Senegal NATIONAL: Mali
Note: These figures combine all the varieties of Arabic. Some data sources, e.g. CIA World Fact Book, World Almanac, Ethnologue, treat these varieties as separate languages.		
6. Portuguese	NATIVE: 203 million 2nd: 10 million TOTAL: 213 million	OFFICIAL: Angola, Brazil, Cape Verde, East Timor, Guinea-Bissau, Macau (People's Republic of China), Mozambique, Portugal, São Tomé e Príncipe.
7. Bengali	NATIVE: 196 million TOTAL: 215 million	OFFICIAL: Bangladesh, India (Tripura, West Bengal)
8. Russian	NATIVE: 145 million 2nd: 110 million TOTAL: 255 million	OFFICIAL: Abkhazia (part of Georgia), Belarus, Kazakhstan, Kyyrgyzstan, Russia, Transnistria (part of Moldova).
9. Japanese	NATIVE: 126 million 2nd: 1 million TOTAL: 127 million	OFFICIAL: Japan, Palau
10. German	NATIVE: 101 million 2nd: 128 million TOTAL: 129 million	OFFICIAL: Austria, Belgium, Germany, Italy (South Tyrol), Liechtenstein, Luxembourg, Poland, Siwtzerland

Intensive English Courses

Live and learn in England. Total immersion in English.

Education Online

Degree-Internet-Kids-Adults Easy enough for an Adult to learn

Ads by Goooooo

Language	# of speakers	Where is it spoken as an official language?
11. Panjabi	Western: 60 million	OFFICIAL: India (Punjab) NATIONAL: Pakistan

219

	Eastern: 28 million TOTAL: 88 million	
12. Javanese	76 million	OFFICIAL: Indonesia (esp. Java)
13. Korean	71 million	OFFICIAL: North Korea, South Korea
14. Vietnamese	NATIVE: 70 million 2nd: 16 million TOTAL: 86 million	OFFICIAL: Vietnam
15. Telugu	NATIVE: 70 million 2nd: 5 million TOTAL: 75 million	OFFICIAL: India (Andhra Pradesh)
16. Marathi	NATIVE: 68 million 2nd: 3 million TOTAL: 71 million	OFFICIAL: India (Daman and Diu, Goa, Maharashtra)
17. Tamil	NATIVE: 68 million 2nd: 9 million TOTAL: 77 million	OFFICIAL: India (Tamil Nadu), Singapore, Sri Lanka
18. French	NATIVE: 67 million 2nd: 63 million TOTAL: 130 million	OFFICIAL or NATIONAL: Belgium, Benin, Burkina Faso, Burundi, Cameroon, Canada, Central African Republic, Chad, Comoros, Congo-Brazzaville, Congo-Kinshasa, Côte d'Ivoire, Djibouti, Equatorial Guinea, France, French Polynesia, Gabon, Guernsey, Guinea, Haiti, India (Karikal, Pondicherry), Italy, Jersey, Lebanon, Luxembourg, Madagascar, Mali, Martinique, Mauritius, Mayotte, Monaco, New Caledonia, Niger, Rwanda, Senegal, Seychelles, Switzerland, Togo, United States (Louisiana), Vanuatu.
19. Urdu	NATIVE: 61 million 2nd: 43 million TOTAL: 104 million	OFFICIAL: India (Jammu and Kashmir), Pakistan.
20. Italian	61 million	OFFICIAL: Croatia (Istria Country), Italy, San Marino, Slovenia, Switzerland.

30 Languages Available

Tap your innate language-learning ability. The Fastest Way to Learn

rosetastone.com

[Ads by Google](#)

Language	# of speakers	Where is it spoken as an official language?
21. Turkish	NATIVE: 60 million 2nd: 15 million TOTAL: 75 million	OFFICIAL: Bulgaria, Cyprus, Turkish Republic of Northern Cyprus, Turkey
22. Persian	54 million	OFFICIAL: Afghanistan, Iran, Tajikistan
23. Gujarati	46 million	OFFICIAL: India (Gujarat, Daman and Diu, Dadra and Nagar Haveli).

220

24. Polish	46 million	OFFICIAL: Poland
25. Ukrainian	39 million	OFFICIAL: Ukraine, Transnistria (part of Moldova).
26. Malayalam	37 million	OFFICIAL: India (Kerala, Lakshadweep, Mahe).
27. Kannada	NATIVE: 35 million 2nd: 9 million TOTAL: 44 million	OFFICIAL: India (Karnataka).
28. Oriya	32 million	OFFICIAL: India (Orissa).
29. Burmese	NATIVE: 32 million 2nd: 10 million TOTAL: 42 million	OFFICIAL: Myanmar.
30. Thai	NATIVE: 20 million 2nd: 40 million TOTAL: 60 million	OFFICIAL: Thailand.

[Language statistics & facts](#) | [Top 30 languages](#) | [20 countries with the most languages](#)

[Language families: By speakers](#) - [By languages](#) | [Languages in the United States](#)

VISTAWIDE

All content on this site is copyrighted. © 2004-2006 VISTAWIDE.COM
[Contact](#) - [About us](#)