

# การวิจัยเพื่อพัฒนาข้อทดสอบสมิทภาพทั่วไปทางภาษาอังกฤษ

อัจฉรา วงศ์โสธร

ศิริพร พงษ์สุรพิพัฒน์

## ความเป็นมาและความสำคัญของปัญหา

จุฬาลงกรณ์มหาวิทยาลัยได้เล็งเห็นว่าภาษาอังกฤษเป็นภาษาสากลในการสื่อสารและแสวงหาความรู้ ตลอดจนมีความสำคัญมากทั้งต่อการศึกษาในระดับอุดมศึกษาและในการประกอบอาชีพในวงการต่างๆ ทั้งภายในและภายนอกประเทศ ถ้าบัณฑิตของจุฬาลงกรณ์มหาวิทยาลัยมีความสามารถในการใช้ภาษาอังกฤษในด้านเนื้อหาวิชาการ และในระดับที่สามารถติดต่อทำความเข้าใจกับต่างประเทศได้ก็จะทำให้บัณฑิตทันต่อเหตุการณ์ มีคุณค่าเป็นที่ต้องการของตลาดแรงงานและสังคมมากขึ้น จุฬาลงกรณ์มหาวิทยาลัยจึงได้จัดโครงการการสร้างและพัฒนาคุณภาพของแบบทดสอบวัดสมิทภาพทั่วไปทางภาษาอังกฤษ และระบบการทดสอบที่เกี่ยวข้องขึ้น เพื่อกระตุ้นให้เกิดการตื่นตัวทั้งอาจารย์และนิสิตให้เกิดความสนใจและให้ความสำคัญต่อการเรียนการสอนภาษาอังกฤษ โดยโครงการนี้มีระยะดำเนินการ 3 ปี คือ ปี 2533-2535 โครงการนี้ได้ดำเนินการเสร็จสิ้นระยะปีที่ 1 ใน พ.ศ. 2533 โดยมีวัตถุประสงค์เพื่อศึกษาสำรวจเกี่ยวกับแนวโน้มระดับสมิทภาพ ทั่วไปทางภาษาอังกฤษของนิสิตชั้นปีที่ 1-6 บัณฑิตวิทยาลัย และบุคคลภายนอกที่สนใจ เพื่อจำแนกออกเป็นมิติต่างๆ คือ เพศ ชั้นปี คณะวิชา เชื้อชาติ สัญชาติ ศาสนา อายุ และรายได้ของครอบครัว ในการดำเนินการในระยะปีที่ 2 คณะผู้วิจัยได้ดำเนินการตามเป้าหมายของโครงการและจากข้อเสนอแนะของคณะผู้วิจัยในระยะปีที่ 1 และได้ดำเนินการเพื่อสร้างและพัฒนาคุณภาพของแบบทดสอบสมิทภาพทั่วไปทางภาษาอังกฤษที่ใช้แล้วในการวิจัยในระยะที่ 1 ให้มีจำนวนข้อเพิ่มมากขึ้นและจัดเก็บในคลังข้อสอบโดยใช้ระบบคอมพิวเตอร์

## วัตถุประสงค์ของการวิจัย

1. เพื่อพัฒนาข้อทดสอบสมิทภาพทั่วไปทางภาษาอังกฤษด้านการฟัง การอ่านและเขียนเพิ่มมากขึ้นจากข้อทดสอบที่ได้จากการสร้างขึ้นในการวิจัยระยะปีที่ 1 และให้ได้แบบทดสอบ 4 ฉบับที่คู่ขนานกันในด้านเนื้อหาและสถิติ
2. เพื่อวิเคราะห์และปรับปรุงแบบทดสอบที่สร้างขึ้นทั้งฉบับและเป็นรายชื่อ

3. เพื่อพัฒนาแบบทดสอบให้เป็นมาตรฐานและเก็บไว้ในธนาคารข้อทดสอบโดยใช้คอมพิวเตอร์ (Computerized item bank)

### ขอบเขตของการวิจัย

1. การสร้างแบบทดสอบสมิททิภาพทั่วไปทางภาษาอังกฤษครั้งนี้สร้างขึ้นเพื่อวัดความรู้ความสามารถทั่วไปทางภาษาอังกฤษเฉพาะการวัดทักษะการฟัง การอ่าน การเขียน เท่านั้น

2. การวิจัยครั้งนี้ผู้วิจัยไม่คำนึงถึงความแตกต่างด้านเพศ อายุ เชาว์นปัญญา ตลอดจนภูมิหลังของผู้เข้าทดสอบ

3. กลุ่มพลวิจัยของการวิจัยครั้งนี้มี 2 กลุ่ม คือ

3.1 พลวิจัยกลุ่มที่ 1 คือ นิสิตจุฬาลงกรณ์มหาวิทยาลัยตั้งแต่ระดับชั้นปีที่ 1 ถึงบัณฑิตศึกษา รวมทั้งบุคคลภายนอกที่สมัครสอบตามประกาศของสถาบัน-ภาษา และเข้าสอบในภาคปลายปีการศึกษา 2533 เมื่อวันที่ 16-31 มกราคม 2534 จำนวน 481 คน

3.2 พลวิจัยกลุ่มที่ 2 คือ บุคคลภายนอกที่มีพื้นฐานความรู้ระดับปริญญาตรีและปริญญาโท ซึ่งเข้าทดสอบในภาคปลายปีการศึกษา 2534 จำนวน 87 คน

### ข้อตกลงเบื้องต้น

คณะผู้วิจัยได้ใช้ตารางกำหนดลักษณะเฉพาะของข้อสอบ (Test Table of Specifications) ที่คณะผู้วิจัยในโครงการระยะปีที่ 1 ได้จัดทำขึ้น ซึ่งเป็นการจัดทำของคณะผู้เชี่ยวชาญทางด้านการออกข้อสอบภาษาอังกฤษ มีความเชื่อถือได้เนื่องจากตารางนี้ได้ใช้เพื่อสร้างข้อทดสอบสมิททิภาพทั่วไปทางภาษาอังกฤษในระยะปีที่ 1 แล้วมีความเที่ยงสูง ( $KR_{20} = 0.90$ ) มีความตรงเชิงประจักษ์ (Face Validity) และความตรงเชิงเนื้อหา (Content Validity) สูง โดยพิจารณาจากเนื้อหาและความสอดคล้องของข้อทดสอบกับวัตถุประสงค์ของการวัดซึ่งระบุไว้ในตารางดังกล่าว

### คำจำกัดความที่ใช้ในการวิจัย

**ข้อทดสอบสมิททิภาพทั่วไปทางภาษาอังกฤษ** หมายถึง ความสามารถทั่วไปทางภาษา ซึ่งไม่ขึ้นต่อหลักสูตร อาจเรียกได้ว่าเป็นสมรรถวิสัยในการใช้ภาษา (Language Competence) ซึ่งมีความถนัดทางภาษา (Language Aptitude) และสัมฤทธิ์ผลทางการเรียน

**ทักษะการฟัง** หมายถึง ทักษะรับสารซึ่งสื่อด้วยเสียง ได้แก่ความสามารถในการฟังเพื่อความเข้าใจข้อความ ข่าว บทสนทนาและการบรรยาย

**ทักษะการอ่าน** หมายถึง ทักษะรับสารซึ่งสื่อด้วยตัวหนังสือ หรือความสามารถในการอ่านเพื่อความเข้าใจข้อความที่ใช้โวหารแบบต่างๆ ได้แก่ การเล่าเรื่อง การแสดงความคิดเห็นเชิงวิพากษ์วิจารณ์ การบรรยายสถานที่ การโฆษณาจูงใจ และการอธิบาย

**ทักษะการเขียน** หมายถึง ทักษะสื่อสารด้วยตัวหนังสือหรือความสามารถใช้ภาษาได้อย่างถูกต้องทั้งทางศัพท์และไวยากรณ์ ความสามารถเขียนข้อความในระดับอนุเขตตามหัวข้อและกรอบเนื้อความที่กำหนดและกรอบเนื้อหา และการเขียนแบบเสรี เขียนความเรียงตามหัวข้อเรื่องที่กำหนดให้

### ประโยชน์ที่คาดว่าจะได้รับ

1. สถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัยได้แบบทดสอบสมิทธิภาพทั่วไปทางภาษาอังกฤษที่มีมาตรฐาน และเป็นที่ยอมรับทั่วไปเพิ่มขึ้นอีก จำนวน 4 ชุด
2. สถาบันภาษาได้มีคลังข้อสอบทางภาษา ซึ่งใช้ระบบคอมพิวเตอร์ในการจัดเพื่อการนำมาใช้และพัฒนาข้อสอบต่อไป

### วิธีดำเนินการวิจัย

#### การสร้างแบบทดสอบ

สถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย ได้แต่งตั้งคณะกรรมการผู้เชี่ยวชาญทางด้านการออกข้อสอบและวัดผลทางภาษาอังกฤษขึ้นมาคณะหนึ่ง คณะผู้ออกข้อสอบได้ดำเนินการสร้างแบบทดสอบดังนี้

1. ศึกษาตารางกำหนดลักษณะเฉพาะของข้อสอบ (Test table of specifications) ซึ่งได้ใช้ในการสร้างข้อสอบสมิทธิภาพทั่วไปทางภาษาอังกฤษและนำมาวิจัยในปีที่ 1 ของโครงการวิจัยนี้แล้วเห็นว่ามีเหมาะสม
2. คณะผู้สร้างข้อทดสอบสมิทธิภาพทั่วไปทางภาษาอังกฤษได้ดำเนินการสร้างตามตารางวิเคราะห์ดังกล่าว ได้จำนวนข้อกระทงดังนี้
  - 2.1 แบบปรนัย (เลือกคำตอบจากตัวเลือก 4 ตัว) แบ่งเป็นข้อกระทงที่สร้างข้อตามทักษะต่าง ๆ ดังนี้

2.1.1	ทดสอบทักษะการฟังเข้าใจความ	105	ข้อ
2.1.2	ทดสอบทักษะการอ่านเข้าใจความ	204	ข้อ
2.1.3	ทดสอบการเขียน	75	ข้อ
2.2	แบบอัตนัย		
2.2.1	วัดการเขียนระดับอนุเจต	6	ข้อ
2.2.2	การเขียนระดับเรียงความ	3	ข้อ

3. คณะผู้ออกข้อสอบได้นำข้อสอบที่พัฒนาและปรับปรุงแล้วจากการสอบระยะที่ 1 พ.ศ. 2533 คละกับข้อสอบที่สร้างขึ้นตามข้อ 2 แล้วดำเนินการคัดเลือกข้อสอบอีกครั้ง ได้จำนวนข้อ กระทั่งดังนี้

3.1	แบบปรนัย แบ่งเป็นข้อกระโทงที่คัดเลือกไว้ดังนี้		
3.1.1	ทดสอบทักษะการฟังเข้าใจความ	140	ข้อ
3.1.2	ทดสอบทักษะการอ่านเข้าใจความ	240	ข้อ
3.1.3	ทดสอบการเขียน	100	ข้อ
3.2	แบบอัตนัย		
3.2.1	วัดการเขียนระดับอนุเจต	8	ข้อ
3.2.2	การเขียนระดับเรียงความ	4	ข้อ

4. นำข้อทดสอบที่คละกันมาจัดเข้าเป็นฉบับตามตารางกำหนดลักษณะเฉพาะทางของ ข้อสอบให้มีเนื้อหา และความยากง่ายของเนื้อหาที่ตัดสินโดยคณะกรรมการพัฒนาข้อทดสอบว่า ไม่แตกต่างกัน ได้แบบทดสอบสมิทธิภาพทั่วไปทางภาษาอังกฤษ 4 ฉบับ

### ผลวิจัย

คณะผู้วิจัยและผู้สร้างแบบทดสอบได้จัดสร้างแบบทดสอบเป็น 4 ฉบับ ผู้วิจัยจึงจัดแบ่ง ผลวิจัยออกเป็น 4 กลุ่ม ตามกลุ่มคณะต่าง ๆ ตามผลการวิจัยตามโครงการวิจัยและพัฒนาการ สร้างและพัฒนาคุณภาพของแบบทดสอบสมิทธิภาพ ปีที่ 1 (สุพัฒน์ สุขมลสันต์ และคณะ 2533) โดยจัดกลุ่มคณะต่างๆ เป็น 4 กลุ่ม ที่ค่าเฉลี่ยของผลการทดสอบวัดสมิทธิภาพภาษาอังกฤษ ครั้งแรก ไม่แตกต่างกันอย่างมีนัยสำคัญ ( $F = .111$   $df = 1,466$ )

## การบริหารการสอบ

1. การผลิต อัดสำเนาและเย็บเข้าเล่มที่สถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย
2. การประชาสัมพันธ์การสอบ ผ่านที่ประชุมคณบดีเพื่อพิจารณาเกี่ยวกับเวลาสอบ และขอความร่วมมือในการเผยแพร่ ประชาสัมพันธ์ให้กับนิสิตในคณะ นอกจากนี้ยังใช้การตีตประกาศตามคณะและที่ป้ายประกาศของสถาบันภาษาและบัณฑิตวิทยาลัย
3. การสมัครสอบ ผู้สนใจสมัครและกรอกใบสมัครที่สถาบันภาษา
4. การจัดสอบ พลวิชัยเข้าสอบเป็นกลุ่ม ที่เป็นนิสิตเข้าตามกลุ่มคณะ ส่วนที่เป็นบุคคลภายนอกเข้าสอบเป็นกลุ่มตามเวลาที่กำหนด สถานที่สอบอยู่ที่คณะและสถาบันภาษา ห้องสอบเป็นห้องเรียนและห้องประชุม การสอบการฟังให้ผ่านลำโพงขยายเสียง ผู้คุมสอบ เป็นอาจารย์และเจ้าหน้าที่ของคณะและของสถาบันภาษา ซึ่งได้ผ่านการอบรมชี้แจงวิธีการดำเนินการสอบเป็นผู้นำแบบทดสอบและกระดาษคำตอบไปดำเนินการสอบ
5. การตรวจข้อสอบ ข้อทดสอบแบบปรนัยตรวจและวิเคราะห์โดยเครื่องคอมพิวเตอร์ของสถาบันบริการคอมพิวเตอร์ของจุฬาลงกรณ์มหาวิทยาลัย ข้อทดสอบแบบอัตนัยตรวจโดยอาจารย์สถาบันภาษา ฉบับละ 2 คน โดยใช้เกณฑ์การให้คะแนนเกณฑ์เดียวกัน

## แนวคิดของการวิจัยทางการทดสอบภาษา (Language Testing Research)

การวิจัยทางการทดสอบภาษามีความหมายโดย 3 นัย ดังนี้

### นัยที่หนึ่ง

การวิจัยทางการทดสอบภาษา หมายถึงการพัฒนาข้อทดสอบทางภาษาขึ้นมาให้เป็นมาตรฐาน โดยนักวิจัยใช้ขั้นตอนดำเนินงานดังนี้

- (1) ศึกษาเกี่ยวกับคุณลักษณะหรือภาวะสันนิษฐาน (Construct) ที่จะวัด และกำหนดว่าประกอบด้วยอะไร
- (2) สร้างเครื่องมือหรือข้อทดสอบที่มีความเที่ยง (Reliability) และความตรง (Validity) ขึ้น เพื่อวัดคุณลักษณะที่ประสงค์
- (3) นำข้อทดสอบไปทดลองใช้และวิเคราะห์คุณภาพของรายข้อและของทั้งฉบับ
- (4) ปรับปรุงข้อทดสอบโดยใช้สถิติวิเคราะห์ที่ได้จากการทดลองข้อทดสอบ
- (5) นำข้อทดสอบไปทดลองใช้อีกครั้งหนึ่ง แล้วตรวจสอบค่าความยากง่าย ค่าความเที่ยง และค่าความตรง

### นัยที่สอง

การศึกษาเกี่ยวกับระดับความสามารถหรือสัมฤทธิ์ผลทางการใช้ภาษา โดยนำเครื่องมือที่เป็นแบบทดสอบทางภาษาไปใช้ในการเก็บข้อมูล

### นัยที่สาม

การศึกษาเกี่ยวกับองค์ประกอบหรือปัจจัย (Factor) ของทักษะ ความสามารถทางการใช้ภาษาว่ามีลักษณะเช่นใด เช่น งานวิจัยของ Oller และ Hinofotis (1980) J. B. Carroll (1983) Bachman (1990:87) และงานวิจัยของ Wongsothorn (1999) ซึ่งให้ข้อมูลเกี่ยวกับปัจจัยที่ประกอบขึ้นมาเป็นสมรรถภาพทางภาษา

โมเดลของ Bachman นั้นมีความคล้ายกับโมเดลทางทฤษฎี ซึ่ง De Saussure (1916) ได้เสนอไว้ว่าประกอบด้วยความสามารถใช้กฎเกณฑ์ภาษา (Langue) และความสามารถใช้ภาษาเพื่อการสื่อความหมาย (Parole) โดยที่ Langue คือ Organizational Competence ตามโมเดลของ De Saussure และ Parole คือ Pragmatic Competence ตามโมเดลของ Bachman

การสร้างข้อทดสอบวัดสมรรถภาพทางภาษาที่สมบูรณ์ต้องครอบคลุมสมรรถภาพทั้ง 2 ด้าน ดังนั้นการวัดจึงควรมุ่งไปยังการใช้ทักษะทางภาษา ได้แก่ การฟัง การพูด การอ่าน การเขียนตามสถานการณ์ในชีวิตจริง ซึ่งครอบคลุมการวัดเนื้อหาของภาษา ได้แก่ การใช้ศัพท์ เสียง โครงสร้าง ความสัมพันธ์ของเนื้อความและศิลปภาษา

### การพัฒนาแบบทดสอบวัดสมรรถภาพทางภาษาแบบอิงเกณฑ์

แบบทดสอบอิงเกณฑ์มักจะใช้กับการสอบสัมฤทธิ์ผล (Achievement Test) ซึ่งอิงหลักสูตร เพราะเป็นการวัดตามวัตถุประสงค์ที่กำหนดไว้ชัดเจนและมีการกำหนดเกณฑ์การผ่านไว้ ส่วนการทดสอบอิงกลุ่มนั้น พัฒนาโดยคำนึงถึงตัวแปรเกี่ยวกับผู้เข้าสอบเป็นสำคัญ และมุ่งเปรียบเทียบผลการสอบระหว่างผู้เรียนมากกว่าเป็นการทดสอบตามวัตถุประสงค์ และเกณฑ์การให้คะแนนก็พิจารณาจากค่าเฉลี่ยและค่าการกระจายของคะแนน

Bachman (1986 : 264) ได้นำเสนอแนวคิดในการพัฒนาแบบทดสอบวัดสมรรถภาพทางภาษาแบบอิงเกณฑ์ว่าเป็นการวัดที่มีประสิทธิภาพ มีอำนาจในการวัดดีกว่าการวัดแบบอิงกลุ่ม เนื่องจากมีการกำหนดวัตถุประสงค์ ขอบเขตเนื้อหาไว้ชัดเจน ข้อกำหนด 2 ประการที่จำเป็นสำหรับการพัฒนาแบบทดสอบวัดสมรรถภาพทางภาษาแบบอิงเกณฑ์ คือ

(1) ข้อทดสอบที่วัดวัตถุประสงค์ตามเนื้อหาซึ่งสุ่มมา ซึ่งเป็นตัวแทนที่แท้จริงของเนื้อหาทั้งหมดที่อยู่ในขอบเขตของเนื้อหา

(2) การกำหนดจุดตัดของคะแนน (Cut-off Score)

Brindley (1986) ได้กล่าวถึงการพัฒนาแบบทดสอบวัดสมรรถภาพทางภาษาแบบอิงเกณฑ์ว่ามุ่งวัดกิจกรรมที่เป็นตัวแทนของการใช้ภาษาที่แท้จริงในชีวิตประจำวัน และการประเมินระดับความสามารถทางการใช้ภาษานั้นจะอิงเกณฑ์ซึ่งเป็นคำอธิบาย (Descriptor) ระดับความสามารถ (Band Scales) ซึ่งมีใช้ต่างๆ กันมากมาย

Bachman (1990 : 256) ได้กล่าวถึงข้อจำกัดของการอธิบายระดับความสามารถและความยากของการกำหนดระดับจากคะแนนที่ได้ และได้เสนอว่าวิธีการเดียวที่จะใช้สำหรับการวัดสมรรถภาพทางการสื่อสารทางภาษา (Communicative Language Competence) ก็โดยการระบุองค์ประกอบของสมรรถภาพและกำหนดระดับความสามารถที่ไม่ขึ้นต่อบริบทใดบริบทหนึ่ง แต่ขึ้นกับความสามารถหรือพฤติกรรมทางภาษาของแต่ละบุคคลอย่างแท้จริง

Bachman (1985) ได้ชี้ให้เห็นว่าการพัฒนาข้อทดสอบนั้นใช้การวิเคราะห์และปรับปรุงเป็นพื้นฐาน และค่าสถิติที่ใช้ได้แก่ค่าความยากง่ายตามแบบทดสอบอิงเกณฑ์นั้น ควรกำหนดตามเกณฑ์ความสามารถ และค่าอำนาจจำแนกจะไม่เน้นเป็นรายข้อ แต่พิจารณาจากค่าเฉลี่ยของฉบับ การกำหนดค่าความยากง่ายจากเกณฑ์ .25-.75 หรือ .20-.80 นั้น เป็นจากเกณฑ์ความสามารถขั้นสูงที่ทำข้อสอบยาก (ค่าความยากง่าย .20 หรือ .25) ได้ จนถึงความสามารถขั้นต่ำ ส่วนที่ทำข้อสอบที่ไม่ยาก (ค่าความยากง่ายค่อนข้าง .80) ได้ ค่าอำนาจจำแนกเฉลี่ยทั้งฉบับนั้นอย่างต่ำควรเป็น .25 - .30

## **การวิเคราะห์ข้อทดสอบแบบอิงกลุ่ม – อิงเกณฑ์ และการวิเคราะห์ตามทฤษฎีความสามารถแฝงแบบ Rasch Model**

### **การวิเคราะห์ข้อทดสอบแบบอิงกลุ่มและแบบอิงเกณฑ์**

ศาสตร์ของการทดสอบทางภาษาในยุคปัจจุบันเป็นยุคของภาษาศาสตร์เชิงจิตวิทยาและสังคมวิทยาและเชิงการสื่อสาร ซึ่ง Morrow (1981) เรียกว่าเป็นยุค The Promised Land การทดสอบหันมาเน้นความสมจริงตามสถานการณ์ที่ใช้ภาษาเพื่อการสื่อสารจริงๆ ในชีวิตประจำวัน ทิศทางของการวิจัยทางการทดสอบภาษาก็ได้หันเหไปจากการเน้นตัวค่าสถิติวิเคราะห์แบบอิงกลุ่ม เพราะสถิติวิเคราะห์ดังกล่าวยึดมั่นอยู่กับกลุ่มตัวอย่าง (Sample Group) ค่าสถิติจึงเปลี่ยนแปลงไปตามกลุ่มที่ทดสอบและการกำหนดมาตรฐานของข้อทดสอบคือ ค่าความยากง่ายและค่าอำนาจจำแนกแบบ อิงกลุ่ม ตลอดจนการเทียบมาตรฐานของแบบทดสอบกับเกณฑ์ภายนอก ได้แก่ การหาความตรงร่วมสมัย (Concurrent Validity) โดยหาค่าสหสัมพันธ์ระหว่างคะแนนที่ผู้สอบได้จาก

แบบทดสอบที่สร้างขึ้นกับคะแนนที่ได้จากแบบทดสอบมาตรฐาน เพื่อพิสูจน์ความเป็นมาตรฐานของแบบทดสอบที่สร้างขึ้น ถูกแทนที่โดยการวิเคราะห์แบบอิงเกณฑ์ และการหาค่าความตรงเชิงเนื้อหา (Content Validity) และเชิงภาวะสันนิษฐาน (Construct Validity) แทน

Pilliner (1983 : 3-4) กล่าวว่า การทดสอบซึ่งมีหลายวัตถุประสงค์ครอบคลุมเนื้อหา มากแต่อยู่ในขอบเขตที่ระบุไว้ชัดเจนโดยใช้กรอบอ้างอิง (Frame of Reference) ที่แน่นอนในการกำหนดเนื้อหาของข้อทดสอบควรเป็นข้อทดสอบอิงเกณฑ์ และใช้ผลการตัดสินคะแนนในแบบต่อเนื่อง (Continuum Model) คือมีระดับได้-ตก มากกว่า 2 ระดับ แทนการใช้วิธีการตัดสินคะแนนที่ใช้เกณฑ์ได้-ตก ที่กำหนดไว้ตายตัวของแบบ State Model ซึ่งแบบหลังนี้เหมาะที่จะใช้ในการสอบวัดความก้าวหน้าระหว่างเรียน (Formative Test) ที่มีวัตถุประสงค์และเนื้อหาบ่อย นอกจากนี้ การทดสอบการสื่อความหมายทางภาษา (Productive) ซึ่งรวมถึงการรับสาร (Receptive) ด้วย จำเป็นต้องครอบคลุมภาวะสันนิษฐาน (Construct) ที่หลากหลายเพื่อวัดการใช้ภาษาที่สมจริงตามสภาพแวดล้อมทางสังคม เนื้อหาและทักษะที่วัดโดยแบบทดสอบจึงหลากหลาย (Heterogeneous) มากกว่าเป็นอันหนึ่งอันเดียว (Homogeneous) ดังนั้นการวิเคราะห์ข้อสอบจึงไม่สามารถใช้วิธีการแบบประเพณีนิยม (Classical Model) ซึ่งการวิเคราะห์ตั้งอยู่บนรากฐานของความเป็นอันหนึ่งอันเดียวกันอย่างเป็นทางการเป็นกลุ่มก้อนหรือมีความเป็นเอกพันธ์ (Homogeneity)

การวิเคราะห์แบบทดสอบทั้งฉบับโดยความเที่ยง (Reliability) ของแบบทดสอบจึงไม่ควรใช้วิธีการหาค่าความมั่นคงภายใน (Internal Consistency) เช่น Kuder Richardson Formula 20 และ 21 หรือ Split-half Pilliner (1983 : 14) ได้แนะนำวิธีการหาค่าความเที่ยงของแบบทดสอบอิงเกณฑ์ว่าควรใช้วิธีการหาค่าสัมประสิทธิ์สหสัมพันธ์แบบ Phi (Phi Coefficient) ค่าสัมประสิทธิ์นี้บางครั้งเรียกว่าค่าสัมประสิทธิ์การอ้างอิงสรุปสำหรับความผิดพลาดที่สมบูรณ์ (Generalizability for absolute error (Brown, 1990 : 16-17) และสามารถใช้อธิบายปริเขต (Domain) หรือขอบเขตเนื้อหาทักษะของการทดสอบ เมื่อข้อสอบที่วิเคราะห์ได้สุ่มอย่างเป็นทางการเป็นตัวแทนของปริเขต สูตรการวิเคราะห์ Phi นี้พัฒนามาจากสูตรของ Brennan ใน Brown (1989) ดังนี้

$$\Phi = \frac{nS_p^2}{n-1} \quad KR_{20} + \frac{x_p(1-x_p) - S_p^2}{k-1}$$



$n$	=	จำนวนผู้เข้าสอบ
$k$	=	จำนวนข้อกระทง
$x_p$	=	ค่ามัธยิมเลขคณิตของคะแนนสัดส่วน (Proportion Score)
$S_p$	=	ค่าเบี่ยงเบนมาตรฐานของคะแนนสัดส่วน (ได้จากสูตรที่ใช้ $N$ คิดไม่ใช่ $N-1$ )
$KR_{20}$	=	ค่าความเที่ยง Kuder Richardson Formula 20
$X_p$	=	$\frac{X}{k}$
$S_p$	=	$\frac{SD}{k}$

สูตรดังกล่าวใช้วิเคราะห์แบบทดสอบอิงเกณฑ์โดยใช้กับคะแนนจากการสอบเพียงครั้งเดียว

การหาค่าความตรง (Validity) ของแบบทดสอบนั้น Pilliner (1983, 14-15) แนะนำวิธีการวิเคราะห์ค่าความตรงเชิงเนื้อหา (Content Validity) โดยให้ผู้เชี่ยวชาญด้านเนื้อหาที่วัดมากกว่า 1 คน ประเมินความสอดคล้องของข้อทดสอบกับเนื้อหา/ทักษะ

นอกจากนี้ Cronbach (1971 อ้างถึงใน Pilliner (1983 : 15) กล่าวถึงขั้นตอนการสร้างแบบทดสอบให้มีความเที่ยงและความตรงว่าเริ่มตั้งแต่การที่คณะผู้สร้างข้อทดสอบเขียนข้อทดสอบจากตารางวิเคราะห์มีความคู่ขนานกันจริงทางสถิติ โดยให้ผู้สอบกลุ่มเดียวกันทดลองสอบทั้ง 2 ฉบับ คะแนนที่ได้ถ้าอยู่ในพิสัยความคลาดเคลื่อนเดียวกัน ถือว่าเป็นคะแนนที่เท่ากัน ความเที่ยงที่หาได้โดยวิธีนี้เป็นค่าความเที่ยงแบบ Test-retest Equivalent Form ส่วนค่าความตรงเป็นค่าความตรงเชิงเนื้อหาเพราะข้อทดสอบพัฒนาจากตารางวิเคราะห์เนื้อหาเดียวกัน

การวิเคราะห์ข้อทดสอบเป็นรายข้อ (Test-item Analysis) สำหรับแบบทดสอบอิงเกณฑ์นั้นจะพิจารณาถึงความสำคัญแก่ค่าความยากง่าย (Difficulty หรือ Facility Index) เท่านั้น ส่วนค่าอำนาจจำแนกไม่มีความสำคัญ ซึ่งต่างกับแบบทดสอบอิงกลุ่มที่ให้ความสำคัญต่อค่าความยากง่ายเท่าเทียมกับค่าอำนาจจำแนก ทั้งนี้เนื่องจากข้อทดสอบอิงเกณฑ์ไม่ได้สร้างขึ้นมาจากอำนาจผู้สอบ (Pilliner, 1983 : 12)

### การวิเคราะห์ข้อสอบตามทฤษฎีความสามารถแฝงแบบ Rasch Model

การวิเคราะห์ข้อทดสอบโดยใช้ Classical Test Theory CTT นั้นอาศัยทฤษฎีที่ไม่มีข้อตกลงเบื้องต้น (Assumption) เกี่ยวกับระดับความสามารถของแต่ละบุคคลกับผลของการทดสอบ อีกทั้งข้อมูลที่ใช้ในการทำนายความสามารถของบุคคลในการทำสอบคือ  $\beta$  นั้น เป็นอัตราส่วนที่คนทั้งกลุ่มทำข้อทดสอบถูกและใช้พาดพิงถึงความสามารถของแต่ละบุคคลในการตอบข้อสอบข้อ

ใดข้อหนึ่งได้โดยเฉลี่ยเท่านั้น จึงจำเป็นต้องมีการวิเคราะห์ที่สัมพันธ์ถึงความสามารถของบุคคลว่ามีความสามารถในระดับใดจึงจะทำข้อสอบได้ ความสัมพันธ์ของระดับความยากง่ายของข้อสอบกับระดับความสามารถของบุคคลในการทำข้อสอบได้ นำไปสู่การวิเคราะห์ข้อสอบตามทฤษฎีความสามารถแฝง (Latent Trait Model) หรือทฤษฎีการสนองตอบต่อข้อกระทง (Item Response Theory/IRT) (Rasch, 1960, Wright and Store, 1979)

IRT มีรากฐานบนข้อตกลงเบื้องต้นที่มั่นคงและเฉพาะเจาะจงกว่า CTT เพราะสามารถทำนายเกี่ยวกับระดับความสามารถของบุคคลในการทำสอบที่มีความยากง่ายในระดับใดระดับหนึ่งได้ IRT มีข้อตกลงเบื้องต้นว่า (1) คุณลักษณะที่วัดเป็นมิติเดียว (Unidimensionality) (2) มาตรการที่ใช้วัดเป็นมิติเดียวและ (3) ความเป็นอิสระในตำแหน่งที่อยู่ของข้อกระทง (Local independence)

การคัดเลือกข้อกระทงแบบ IRT นั้น จะตัดข้อที่ไม่อยู่ในโค้งลักษณะข้อกระทง (Item Characteristic Curve/ICC) ICC เป็นรากฐานของ IRT (Bachman, 1990 : 205-206) เพราะกำหนดข้อตกลงเบื้องต้นเกี่ยวกับความสัมพันธ์ของความเป็นไปได้ของความสามารถของบุคคลกับความเป็นไปได้ที่จะทำข้อกระทงได้ เกณฑ์ของการกำหนด ICC เต็มรูปแบบประกอบด้วย

- (1) ค่าอำนาจจำแนกของข้อทดสอบ (a)
- (2) ค่าความยากง่ายของข้อทดสอบ (b)
- (3) การเดาถูกของข้อทดสอบ (c)

ในการวิเคราะห์ข้อสอบโมเดล IRT นี้ อาจทำได้โดยใช้เกณฑ์ a และ b คือ Two parameter ซึ่งมีข้อตกลงเบื้องต้นว่าคนที่มีความสามารถต่ำจะไม่สามารถตอบข้อสอบถูก และค่าการเดา =  $\theta$  ในการทดสอบภาษาโมเดลที่นิยมใช้กันมากเป็นแบบพารามิเตอร์เดียวหรือ One parameter ของ Rasch ซึ่งกำหนดให้ค่าอำนาจจำแนกของข้อกระทงทุกข้อเท่ากัน และไม่มีค่าการเดา มีการใช้ Rasch อย่างแพร่หลายและอย่างมีประสิทธิภาพในการวิเคราะห์ข้อกระทง เช่นกล่าวถึงใน Davidson และ Henning (1985) Griffin (1985) Henning Hudson and Turner (1985) Madsen และ Larson (1986) Pollitt และ Hutchinson (1987) Larson (1987) Madsen (1987) Adams Griffin (1985) Griffin and Martin (1987) และ McNamara (1990) ได้กล่าวถึงการนำ Rasch วิเคราะห์ ข้อทดสอบภาษาเพื่อการสื่อสาร นอกจากนี้ Laurier (1990) ได้กล่าวถึงการนำ IRT โดยระบบคอมพิวเตอร์ขนาดเล็ก (Micro-computer) ในการจัดทำข้อทดสอบจัดระดับภาษาฝรั่งเศส

## การสร้างธนาคารข้อทดสอบ (Test-item Bank)

ข้อทดสอบที่ผ่านการวิเคราะห์ความเป็นมาตรฐานแล้ว ควรนำเข้าเก็บไว้ในธนาคารข้อทดสอบเพื่อการนำมาใช้ต่อไป

(1) การสร้างธนาคารข้อสอบโดยใช้คอมพิวเตอร์ (Computerized item banking) ช่วยในการคัดเลือกข้อสอบใช้แต่ละครั้งและในการจัดพิมพ์ ผลิตข้อสอบ ตลอดจนการพัฒนาข้อสอบขึ้นใช้อย่างต่อเนื่อง โดยคอมพิวเตอร์แบบตั้งโต๊ะที่ใช้ในการจัดระบบตั้งแต่การวิเคราะห์ จัดเก็บ คัดเลือกในการนำออกมาใช้ในแต่ละครั้ง และจัดพิมพ์ข้อทดสอบที่อยู่ในธนาคารต้องมีมากกว่าจำนวนที่จำเป็นที่จะต้องใช้ในแต่ละครั้งอย่างน้อยมากกว่า 1 เท่าขึ้นไป และควรมีปริมาณตามอัตราส่วนของความสำคัญ และความจำเป็นของเนื้อหา ทักษะทางภาษา

(2) การสร้างธนาคารข้อทดสอบสร้างขึ้นตามการใช้ภาษาในชีวิตจริงเหมาะสมกับการจัดเก็บข้อทดสอบวัดสมิทธิภาพทั่วไปทางภาษาแบบอิงเกณฑ์ โดยการใช้ตารางวิเคราะห์ภาษาที่ใช้ในชีวิตประจำวัน โดยนักวิชาการทางภาษาและนักวัดผลทางภาษา ดังตัวอย่างต่อไปนี้คือ

ทักษะ	เนื้อหา	หน้าที่/วัตถุประสงค์	ความสำคัญ
อ่าน	เอกสาร	จับใจความ	25%
	วารสาร	สรุปความ	
	หนังสือพิมพ์		
	บทความ		
	เรื่องสั้น		
	นวนิยาย		
ฟัง	บทสนทนา	เพื่อการติดต่อทางสังคม	25%
	เพลง	เพื่อความรู้ทันโลก	
	ข่าว	เพื่อความบันเทิง	

ทักษะ	เนื้อหา	หน้าที่/วัตถุประสงค์	ความสำคัญ
อ่าน-เขียน	จดหมาย	ถาม/ตอบ เพื่อให้ได้ข้อมูลรายละเอียด	20%
	บทความ	สรุปประเด็น	
	เรื่องสั้น	วิจารณ์	
เขียน	บันทึก	อ้างอิง	
		บันทึกความจำ	
	รายงาน	ให้ข้อมูล	
ฟัง-พูด	เรื่องส่วนตัว/กิจกรรม	ถาม/ตอบ เพื่อให้ได้ข้อมูล	30%
	ได้วาที	เสนอข้อคิดเห็นโต้แย้ง	
พูด	บรรยาย/อธิบาย	ให้ข้อมูลความรู้	

(3) การสร้างธนาคารข้อทดสอบสำหรับการวัดความสามารถในการใช้ทักษะรับสารทางภาษา ได้แก่ การฟัง อ่าน และทักษะสื่อสารทางภาษา ได้แก่ พูด (จัดเก็บตัวคำถาม บทบาทที่กำหนด สถานการณ์การพูด) และการเขียน (จัดเก็บ คำสั่ง หัวข้อการเขียน)

(4) การสร้างธนาคารข้อทดสอบให้ประโยชน์ทั้งทางด้านบริหารและด้านวิชาการ

ประโยชน์ทางด้านบริหาร คือช่วยไม่ให้เกิดการซ้ำซ้อนกันในการทำงาน การสิ้นเปลืองเวลาและทรัพยากรที่ต้องเสียไปในการสร้าง พัฒนา และใช้ข้อทดสอบที่สร้างขึ้นเพื่อใช้ตามวัตถุประสงค์เดียวกันเพราะธนาคารข้อทดสอบเป็นแหล่งรวบรวมข้อสอบภาษาไว้ตามพารามิเตอร์ประโยชน์ทางด้านวิชาการ มีดังนี้

(1) ทำให้รักษามาตรฐานของการทดสอบและประเมินผลทางภาษา

(2) ทำให้มีข้อทดสอบที่ได้มาตรฐานและสร้างตามพารามิเตอร์ที่ฟังประสงค์ได้

เป็นจำนวนมากมาย

(3) ทำให้การทดสอบเป็นไปตามเอกัตภาพ (Individualized Testing) วัดความ

สามารถของบุคคลตามเกณฑ์และจัดสอบได้เป็นรายบุคคล

- (4) ทำให้การทดสอบไม่จำกัดระดับโดยมีข้อทดสอบที่มีระดับความยากง่ายต่างๆ กัน
- (5) ทำให้มีการพัฒนาข้อทดสอบให้เหมาะสมกับความสามารถของผู้เรียนแต่ละคน (Tailored Testing)
- (6) ทำให้เพิ่มพูนศักยภาพในการเรียนรู้ภาษาและสมรรถวิสัยในการใช้ภาษาเพื่อการสื่อสารของผู้เรียน โดยการวัดความสามารถของตนอย่างสม่ำเสมอด้วยข้อสอบที่เป็นมาตรฐาน
- (7) ทำให้ผู้สอบสามารถฟื้นฟูปรับปรุงทักษะภาษาของตนได้อย่างสม่ำเสมอ โดยการวัดผลที่ต่อเนื่องแบบเอกัตบุคคล

## สรุป

วิธีวิเคราะห์ข้อทดสอบที่ใช้ในการวิจัยนี้ มี 2 ขั้นตอน คือ โดยวิธี Classical Model และ Rasch Model ตามลำดับ

1. Classical Model ให้ข้อมูลที่เป็นค่าความยากง่าย และอำนาจจำแนกของแต่ละตัวเลือก นำมาวิเคราะห์ตัวเลือกที่ใช้ทุกตัว มีผลเพื่อการปรับปรุงข้อทดสอบเป็นรายข้อ
2. Rasch Model ใช้เพื่อคัดเลือกข้อทดสอบที่มีคุณสมบัติสอดคล้องกับโครงสร้างลักษณะของข้อทดสอบ (Item Characteristic Curve / ICC) สำหรับจัดเป็นธนาคารข้อทดสอบ ดังนั้นธนาคารจะบรรจุข้อทดสอบมาตรฐานที่ผ่านการปรับปรุงเป็นรายข้อจากผลการวิเคราะห์ Classical Model อีกทั้งเป็นข้อทดสอบที่สอดคล้องกับ ICC จึงเป็นการยืนยันยืนยันความเป็นมาตรฐานของข้อทดสอบในธนาคารได้

## ผลของการวิจัย

สรุปได้ดังนี้

1. ได้แบบทดสอบปรนัยวัดทักษะการฟัง อ่าน และเขียน และแบบอัตนัยวัดทักษะการเขียน จำนวน 4 ฉบับ ที่มีค่าความเที่ยงอยู่ในเกณฑ์ ดี - ดีมาก ( $KR_{20} = .854 - .910$  แบบ Phi =  $.822 - .896$ ) และแบบอัตนัยวัดทักษะการเขียนที่มีค่าความเที่ยงอยู่ในเกณฑ์ ดี - ดีมาก ( $R_{xy} = .594 - .840$ ) โดยมีค่าอำนาจในการทำนาย  $48.00 - 58.50$
2. ได้ข้อทดสอบแบบปรนัยวัดสมิทธิภาพทั่วไปทางภาษาอังกฤษที่มีคุณภาพดีมาก - พอใช้ได้ จำนวน 471 ข้อ
3. ได้ข้อทดสอบแบบอัตนัยวัดสมิทธิภาพทางการเขียนที่มีคุณภาพ ดี - พอใช้ได้ 12 ข้อ

4. ได้ข้อทดสอบแบบปรนัยวัดสมิทริภาพทั่วไปทางภาษาอังกฤษซึ่งได้มาตรฐานเข้าเกณฑ์ของ Rasch Model จำนวน 414 ข้อ

### อภิปรายผล

การวิเคราะห์แบบทดสอบสมิทริภาพทั่วไปทางภาษาอังกฤษแบบปรนัย ซึ่งมีการให้คะแนน 0 และ 1 ใช้วิธีการต่างกับแบบทดสอบอัตนัย ซึ่งมีการให้คะแนนถึง 2 หลัก ทั้งการหาค่าความเที่ยง ค่าความยากง่าย และค่าอำนาจจำแนก และเนื่องจากว่าแบบทดสอบวัดสมิทริภาพทั่วไปทางภาษาอังกฤษนี้เป็นแบบทดสอบอิงเกณฑ์ การวิเคราะห์จึงให้ความสำคัญต่อค่าความยากง่ายมากกว่าค่าอำนาจจำแนก

การวิเคราะห์ข้อทดสอบโดยใช้ Classical Model เหมาะสมกับการปรับปรุงข้อกระทง เพราะให้ข้อมูลเกี่ยวกับตัวเลือกที่ถูกต้องและตัวลวง ส่วนการวิเคราะห์โดยใช้ Rasch Model เหมาะสมสำหรับการคัดเลือกข้อกระทงเพื่อนำไปเก็บไว้ในธนาคารข้อทดสอบเพราะเป็นการวิเคราะห์ที่ไม่อิงกลุ่ม (Sample-free)

การปรับปรุงข้อกระทงโดยยึดค่าความยากง่ายเป็นเกณฑ์และอำนาจจำแนกที่ไม่ดีดลบจะทำให้ค่าอำนาจจำแนกของข้อสอบดีขึ้นตามไปด้วย

### ข้อเสนอแนะ

1. ควรจัดเก็บข้อกระทงที่มีคุณภาพมาตรฐานตามเกณฑ์ไว้ใช้ในธนาคารข้อทดสอบโดยใช้คอมพิวเตอร์

2. ควรพัฒนาข้อทดสอบทักษะภาษาแบบอิงเกณฑ์ที่วัดสมิทริภาพทั่วไปทางภาษาอังกฤษที่ใช้ตารางวิเคราะห์อื่นๆ ซึ่งสร้างขึ้นจากสภาพการณ์ของการใช้ภาษาเพื่อการสื่อสารที่แตกต่างหลากหลายออกไป

3. ควรทำการวิจัยเปรียบเทียบระดับสมิทริภาพทางการใช้ภาษาอังกฤษของนิสิตนักศึกษาและบุคคลทั่วไปที่วัดโดยแบบทดสอบของสถาบันภาษากับแบบทดสอบมาตรฐานที่มีการใช้อย่างแพร่หลาย เช่น แบบทดสอบ TOEFL (Test of English as a Foreign Language) ซึ่ง Educational Testing Service แห่ง Princeton, New Jersey พัฒนาค้นขึ้นใช้และแบบทดสอบ ELTS (English Language Testing Service) ซึ่ง British Council และ University of Cambridge Local Examinations Syndicate พัฒนาค้นขึ้น

## เอกสารอ้างอิง

- โกวิท ประวาลพุกฤษ์ และสมศักดิ์ สินธุเวชญ์. 2527. *การประเมินในชั้นเรียน*. กรุงเทพฯ: วัฒนาพานิช.
- สุพัฒน์ สุขมลสันต์ และคณะ. 2533. รายงานการวิจัยฉบับที่ 1 ในโครงการ การสร้างและพัฒนา คุณภาพของแบบทดสอบวัดสมิทธิภาพทั่วไปทางภาษาอังกฤษและระบบการทดสอบ ที่เกี่ยวข้องเรื่อง “แนวโน้มระดับสมิทธิภาพทั่วไปทางภาษาอังกฤษของนิสิต จุฬาลงกรณ์มหาวิทยาลัยและบุคคลภายนอกที่สนใจ ปี พ.ศ. 2533”. สถาบันภาษา และ ฝ่ายวิชาการ จุฬาลงกรณ์มหาวิทยาลัย.
- อัจฉรา วงศ์ไธธร. 2529. *เทคนิควิธีการสร้างข้อสอบภาษาอังกฤษสำหรับวัดและประเมินผลการใช้ภาษาเพื่อการสื่อสาร*. กรุงเทพฯ: สำนักพิมพ์อักษรเจริญทัศน์.
- \_\_\_\_\_. 2532. การวิจัยเพื่อพัฒนาแบบทดสอบอิงปริเจทภาษาอังกฤษสำหรับใช้กับนักศึกษา ไทยในระดับต่างๆ. งานวิจัยจุฬาลงกรณ์มหาวิทยาลัย ลำดับที่ 21. ฝ่ายวิจัย จุฬาลงกรณ์มหาวิทยาลัย. กรุงเทพฯ: โรงพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.
- Adams, R.J., R.E. Griffin, and L. Martin. 1987. A latent trait method for measuring dimension in second language proficiency. *Language Testing* 4, 1 : 9-27.
- Bachman, L.F. 1985. “Critical review of current practices in language testing and evaluation.” Lecture presented at CULI National Seminar on *Problems and Issues in ELT and ELTE at the Tertiary Level*, Bangkok 15 May-4 June 1985.
- \_\_\_\_\_. 1986. “The development and use of criterion-referenced tests of language proficiency in language program evaluation.” In K. Prapphal et al. (eds.): *Trends in Language Programme Evaluation* (Papers presented at CULI’s First International Conference, 9-11 December, 1986, Bangkok). Bangkok: Chulalongkorn University Language Institute.
- \_\_\_\_\_. And A.S. Palmer. 1982. “The construct validation of some components of communicative proficiency.” *TESOL Quarterly* 16, 4 : 449-65.
- Bachman, L.F. 1990. *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Brindley, G. 1986. *The Assessment of Second Language Proficiency : Issues and Approaches*. Adelaide : National Curriculum Resource.
- Brown, J.D. 1989. “Short-cut estimates of criterion-referenced test reliability.” Paper presented at the 11<sup>th</sup> Annual Language Testing Research Colloquium, San Antonio, March, 1989.

- Carroll, J.B. 1983. "The difficulty of a test and its factor composition revisited" in H. Wainer and S. Messick (eds.): *Principles of Modern Psychological Measurement: A Festschrift for Frederic M. Lord*. Hillsdale, NJ: Lawrence Erlbaum Associates: 257-82.
- Cronbach, L.J. 1951. "Coefficient Alpha and the internal structure of tests". *Psychometrika*, 16, 297-334.
- Davidson, F. and G. Henning. 1985. "A self-rating scale of English proficiency: Rasch scalar analysis of item and rating categories. *Language Testing* 2, 2 : 164-79.
- De Saussure, F. 1916. *Cours de linguistique generale*. Paris: Payot. [Course in general linguistics (translated by Wade Baskin). New York: Philosophical Library, 1985.]
- Griffin, P.E. 1985. The use of latent trait models in the calibration of tests of spoken language in large-scale selection-placement programs." In Y.P. Lee, A.C.Y. Fok, R. Lord and G. Low. (eds.) *New directions in language testing*. Oxford: Pergamon Press.
- Henning, G., T. Hudson, and J. Turner. 1985. "Item response theory and the assumption of unidimensionality." *Language Testing* 2, 2 : 141-54.
- Larson, J.W. 1987. "Computerized adaptive language testing: A Spanish placement exam." In K.M. Bailey, T.L. Dale, and R.T. Clifford. (eds.) 1987. *Language Testing Research: Selected Papers from the 1986 Colloquium*. Monterey, CA: Defense Language Institute, 1-19.
- Laurier, M. 1990. "What we can do with computerized adaptive testing and what we cannot do." Paper presented at SEAMEO RELC 25<sup>th</sup> Regional Seminar on *Language Testing and Language Programme Evaluation*, 9-12 April 1990, Singapore.
- Madsen, H.S. 1987. "Utilizing Rasch analysis to detect cheating on language examinations" in K.M. Bailey, T.L. Dale, and R.T. Clifford (eds.) 1987. *Language Testing Research: Selected Papers from the 1986 Colloquium*. Monterey, CA: Defense Language Institute.
- \_\_\_\_\_. and J.W. Larson. 1986. "Computerized Rasch analysis of item bias in ESL tests" in C.W. Stansfield. (ed.) 1986. *Technology and Language Testing*. (Collected papers from the 1985 Colloquium.) Washington, DC: TESOL, 11-23.
- McNamara, T.F. 1990. "The role of item response theory in language test validation." Paper presented at SEAMEO RELC 25<sup>th</sup> Regional Seminar on *Language Testing and Language Programme Evaluation*, 9-12 April 1990, Singapore.



- Morrow, K. 1981. "Communicative language testing: Revolutions or evolution" in C.J. Alderson and A. Hughes. (eds.) *ELT Documents 111—Issues in Language Testing*. London: The British Council.
- Oller, J.W. Jr. and F.B. Hinofotis. 1980. "Two mutually exclusive hypotheses about second language ability: Indivisible and partly divisible competence. In J.W. Oller, Jr. and K. Perkins (eds.) 1980. *Research in language testing*. Rowley, MA: Newbury House, 13-23.
- \_\_\_\_\_. (Ed.) 1983. *Issues in language testing research*. Rowley, MA: Newbury House.
- Pilliner, A.G. 1983. "Norm-referenced and criterion-referenced tests -- and evaluation." (Paper presented at CULI National Seminar on *Evaluation Procedures for Communicative Syllabuses*, 25 October – 4 November 1983, Bangkok: Chulalongkorn University Language Institute.
- Pollitt, A. and C. Hutchinson. 1987. "Calibrating graded assessments: Rasch partial credit analysis of performance in writing." *Language Testing* 4, 1 : 72-92.
- Rasch, G. 1960. *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danmarks Paeagogiske Institute.
- Wongsothorn, A. 1999. "Factors in English abilities of Thai students: From a perspective of English curriculum design." *Thailand's TESOL Bulletin*. August 1999 : 18-23.
- Wright, B.D. and M.H. Stone. 1979. *Best test design*. Chicago: Mesa Press.