

การพัฒนาแบบทดสอบสมรรถภาพ ทางภาษาอังกฤษเพื่อวิชาชีพ และการสื่อสารนานาชาติของ สถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย และโปรแกรมการทดสอบให้ได้มาตรฐาน

รองศาสตราจารย์ ดร. สุวัฒน์ สุขมลสันต์

ความนำ

ปัจจุบันนี้เป็นยุคของสารสนเทศและเทคโนโลยีเพื่อการสื่อสาร (Information and Communication Technology Age) ที่มีความก้าวหน้าหลาย ๆ ด้าน เช่น วิทยาการคอมพิวเตอร์ การสื่อสาร การทำธุรกิจผ่านทางระบบเครือข่ายคอมพิวเตอร์ และการเรียนรู้ผ่านทางระบบเครือข่ายดังกล่าว เป็นต้น นอกจากนี้คนทั่วโลกนิยมใช้ภาษาอังกฤษเป็นสื่อสากลเพื่อการสื่อสารหรือสืบค้นข้อมูลต่าง ๆ ที่เกี่ยวข้องทั้งทางระบบเครือข่ายคอมพิวเตอร์ หรือสื่ออื่น ๆ มากที่สุดในโลก กล่าวคือ ประมาณร้อยละ 30 ของผู้ใช้ระบบเครือข่ายคอมพิวเตอร์หรือประมาณ 312,924,679 ล้านคนใช้ภาษาอังกฤษ (Internet World Stat, 2006) นอกจากนี้ภาษาอังกฤษยังเป็นภาษาที่ใช้ในสิ่งพิมพ์มากที่สุดในโลก รวมทั้งเป็นภาษาซึ่งผู้ที่ไม่ใช่เจ้าของภาษาใช้เพื่อพูดเป็นจำนวนมากที่สุดในโลกด้วย (Vistawide, 2006) ด้วยเหตุนี้จึงเป็นที่ยอมรับกันทั่วไปได้ว่าภาษาอังกฤษมีความจำเป็นต่อการศึกษาค้นคว้าของนิสิตนักศึกษาและคนทั่วไป รวมทั้งมีความสำคัญต่อการประกอบธุรกิจของนักธุรกิจทั่วไปและการประกอบอาชีพของผู้ที่อยู่ในวงการอื่น ๆ อีก

รองศาสตราจารย์ ดร. สุวัฒน์ สุขมลสันต์

- ก.ศ.บ. (เกียรตินิยม), (อังกฤษ-ชีววิทยา), วิทยาลัยการศึกษามหาสารคาม พ.ศ. 2513
- ค.ม. (การสอนภาษาอังกฤษ), จุฬาลงกรณ์มหาวิทยาลัย พ.ศ. 2517
- Cert. in TEFL/TESL, RELC/Singapore พ.ศ. 2520
- Dip. in Computer Programming, Edinburgh University, Scotland พ.ศ. 2521
- M.Sc. (Applied Linguistics), Edinburgh University, Scotland พ.ศ. 2522
- Cert. in CALL, Lancaster University, England พ.ศ. 2525
- ค.ด. (การวัดและประเมินผลการศึกษา), จุฬาลงกรณ์มหาวิทยาลัย พ.ศ. 2528

มีความชำนาญในด้านการเรียนการสอนภาษาอังกฤษ การทดสอบทางภาษา สถิติเพื่อการวิจัย วิธีวิทยาการวิจัยทางภาษาศาสตร์และการศึกษา การเขียนโปรแกรมคอมพิวเตอร์ และการใช้คอมพิวเตอร์เพื่อการวิจัยและเพื่อช่วยการเรียนการสอน มีผลงานทางวิชาการหลายเล่ม ทั้งเป็นกรรมการผู้ทรงคุณวุฒิของสถาบันการศึกษาต่าง ๆ หลายแห่ง

อนึ่ง มหาวิทยาลัยมีหน้าที่และความรับผิดชอบโดยตรงต่อสังคมในการช่วยสร้างบัณฑิตให้มีความรู้และความสามารถทันต่อการเปลี่ยนแปลงของสังคม และจากการวิจัยเชิงสำรวจจำนวนมากมักพบว่านักศึกษา บัณฑิตและหน่วยงานต่าง ๆ ของราชการ องค์กรของรัฐวิสาหกิจ และบริษัทห้างร้านเอกชนมีความต้องการและมีการใช้ภาษาอังกฤษในการบริหารกิจการในระดับปานกลางถึงในระดับสูง โดยเฉพาะอย่างยิ่งผู้ที่เป็นนักวิชาการและบุคคลในระดับบริหารกิจการต้องใช้ภาษาอังกฤษเพื่อการอ่าน เขียน ฟัง และพูดในระดับมากตามลำดับ ปริมาณความต้องการเช่นนี้มีมานานแล้ว และมีแนวโน้มว่าจะมีมากขึ้นเรื่อย ๆ (อัจฉรา วงศ์โสธร และคณะ, 2523:20; บังอร สว่างวิโรธ, 2527:38 สุพัฒน์ สุขมลสันต์ และคณะ, 2529:63-68)

ดังนั้น ในปี พ.ศ. 2533 จุฬาลงกรณ์มหาวิทยาลัยได้เล็งเห็นความจำเป็นอย่างยิ่งในการสื่อสารระหว่างประเทศ โดยใช้ภาษาอังกฤษเป็นสื่อกลาง โดยเชื่อว่าหากบัณฑิตของจุฬาลงกรณ์มหาวิทยาลัยมีความสามารถทั้งทางเนื้อหาวิชาการ และการใช้ภาษาอังกฤษในระดับที่สามารถติดต่อกับชาวต่างประเทศได้ ก็จะทำให้บัณฑิตมีคุณค่าเป็นที่ต้องการของตลาดแรงงานและสังคมมากขึ้น ซึ่งจะทำให้เป็นแรงผลักดันให้สังคมไปสู่สังคมอุตสาหกรรมได้ทันเวลาและปลอดภัย ดังนั้น ฝ่ายวิชาการของมหาวิทยาลัยจึงเห็นความจำเป็นที่จะต้องปรับความสามารถในการใช้ภาษาอังกฤษของบัณฑิตให้ได้มาตรฐานที่ทางมหาวิทยาลัยกำหนด และวิธีหนึ่งที่จะทำได้ก็คือ การทดสอบสมรรถภาพทั่วไปทางภาษาอังกฤษของนิสิตชั้นปีที่ 3-4 หรือ 2 ปีสุดท้ายของหลักสูตรต่าง ๆ เพื่อเป็นการกระตุ้นให้อาจารย์และนิสิตตื่นตัวและให้ความสำคัญต่อการเรียนการสอนภาษาอังกฤษด้วยวิธีการต่าง ๆ (อัจฉรา วงศ์โสธร และคณะ, 2535; สุพัฒน์ สุขมลสันต์ และคณะ, 2533) ดังนั้นในปี พ.ศ. 2533 ฝ่ายวิชาการของมหาวิทยาลัย จึงได้ร่วมกับฝ่ายวิจัยและพัฒนาของสถาบันภาษา ทำการวิจัยและพัฒนาแบบทดสอบวัดสมรรถภาพทั่วไปทางภาษาอังกฤษของจุฬาลงกรณ์มหาวิทยาลัย (Chulalongkorn University Test of English Proficiency หรือ CU-TEP) ขึ้นตามแนวแบบทดสอบ TOEFL (Test of English as a Foreign Language) โครงการวิจัยนี้เป็นโครงการต่อเนื่องเป็นเวลา 3 ปีระหว่างปี พ.ศ. 2533 ถึง 2536 (สุพัฒน์ สุขมลสันต์ และคณะ, 2536) แต่หลังจากนั้นสถาบันภาษาได้มีการสร้างและพัฒนาแบบทดสอบดังกล่าวติดต่อกันเรื่อยมาเพื่อให้บริการการทดสอบแก่นิสิตนักศึกษา และบุคคลทั่วไปที่สนใจ ทั้งภาครัฐ รัฐวิสาหกิจและเอกชน จนเป็นที่ยอมรับกันทั่วไปในวงการศึกษาและวงการธุรกิจบางสาขา และทำให้มหาวิทยาลัยจัดตั้งศูนย์ทดสอบทางวิชาการแห่งจุฬาลงกรณ์มหาวิทยาลัย (Academic Testing Center of Chulalongkorn University) ขึ้นเพื่อให้บริการดังกล่าวในปัจจุบันนี้

แต่อย่างไรก็ตาม เนื่องจากแบบทดสอบ CU-TEP เป็นแบบทดสอบสำหรับผู้ที่ต้องการวัดความสามารถทั่วไปทางภาษาอังกฤษของตนเอง จึงยังไม่เหมาะสำหรับการทดสอบความสามารถเฉพาะทางเพื่อวิชาชีพและการสื่อสารนานาชาติ ซึ่งเป็นแบบทดสอบที่สังคมมีความต้องการมากขึ้น ดังจะเห็นได้จากมีแบบทดสอบที่นิยมใช้ทั่วไปขณะนี้ คือแบบทดสอบ TOEIC (Test of English for International Communication) ของ ETS (Educational Testing Service) แห่งสหรัฐอเมริกา แต่ค่าบริการทดสอบของแบบทดสอบนี้ยังสูงมากสำหรับคนทั่วไป

ดังนั้น สถาบันภาษาจึงเห็นความสำคัญที่จะจัดให้มีโครงการวิจัยเรื่อง การพัฒนาแบบทดสอบวัดสมรรถภาพทางภาษาอังกฤษเพื่อวิชาชีพและการสื่อสารนานาชาติของสถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย หรือ CULI-TEPOIC (Chulalongkorn University Language Institute Test of English Proficiency for Occupational and International Communication Purposes) เพื่อสร้างแบบทดสอบที่ได้มาตรฐานและทำการพัฒนาโปรแกรมคอมพิวเตอร์ที่จำเป็นต้องใช้เพื่อการพัฒนาข้อทดสอบและการทดสอบให้ทันสมัยยิ่งขึ้น ทั้งนี้เพื่อให้ระบบการทดสอบของสถาบันภาษามีความสมบูรณ์ยิ่งขึ้นด้วย

วัตถุประสงค์ของโครงการวิจัย

1. เพื่อสร้างและพัฒนาแบบทดสอบวัดสมรรถภาพทางภาษาอังกฤษเพื่อวิชาชีพและการสื่อสารนานาชาติของสถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย ให้มีคุณภาพและได้มาตรฐาน จำนวน 3 ชุด

2. เพื่อพัฒนาโปรแกรมคอมพิวเตอร์ที่ใช้ในการพัฒนาแบบทดสอบและการทดสอบให้ทันสมัย 5 โปรแกรม

ขอบข่ายของการวิจัย

1. แบบทดสอบที่สร้างและพัฒนาเป็นแบบทดสอบวัดสมรรถภาพทางภาษาอังกฤษเพื่อวิชาชีพและการสื่อสารนานาชาติจำนวน 3 ชุดๆ ละ 100 ข้อในลักษณะที่คล้ายคลึงกับแบบทดสอบ TOEIC กล่าวคือแบบทดสอบแต่ละชุดประกอบด้วยส่วนต่างๆ ดังนี้

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 (Sentence Completion) 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. แบบทดสอบที่สร้างและพัฒนาเป็นแบบทดสอบที่เน้นเรื่องการสื่อสารเพื่อวิชาชีพและการสื่อสารนานาชาติสำหรับนิสิต/นักศึกษาในระดับมหาวิทยาลัยและบุคคลทั่วไปที่สนใจ
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. ผู้สอบแบบทดสอบที่ใช้ในการวิจัยครั้งนี้ได้พยายามมากที่สุดในการตอบแบบทดสอบ เพื่อให้ได้คะแนนมากที่สุดเท่าที่จะสามารถทำได้ เพราะการทดสอบแต่ละครั้งเป็นการทดสอบจริงเพื่อทดสอบความรู้และความสามารถของตนเองว่ามีมากน้อยเพียงใด หรือเพื่อนำผลการทดสอบไปประกอบการพิจารณาสมัครเข้าทำงานของตนเอง ซึ่งเป็นความต้องการของผู้สอบแต่ละคน
2. โปรแกรมคอมพิวเตอร์ทั้ง 5 โปรแกรมที่ใช้ในระบบ DOS เป็นโปรแกรมที่ทำงานได้ผลการคำนวณถูกต้องและเชื่อถือได้ เนื่องจากเป็นโปรแกรมที่ได้พัฒนาและตรวจสอบผลการทำงานมานานแล้ว ว่าถูกต้องและเชื่อถือได้ (แต่ที่จำเป็นต้องปรับปรุงและพัฒนาให้สามารถทำงานได้ในระบบ 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 Specifications) เดียวกัน แต่ทำการทดสอบกับกลุ่มผู้สอบที่แตกต่างกัน แม้ว่าจะอยู่ในระดับอุดมศึกษาและบุคคลทั่วไป เช่นเดียวกันก็ตาม ดังนั้น ค่าสถิติที่ได้จากการวิเคราะห์ข้อทดสอบแต่ละชุดอาจแตกต่างกันบ้าง
2. การคำนวณหาความตรงร่วมสมัย (Concurrent Validity) ของแบบทดสอบที่สร้างขึ้นทั้ง 3 ชุด อาศัยคะแนนผลการสอบจากแบบทดสอบ TOEIC กับคะแนนผลการสอบจากแบบทดสอบที่สร้างขึ้น ผู้วิจัยได้สุ่มเลือกผู้สอบได้คะแนนในระดับต่างๆ อย่างมีระบบจำนวนชุดละ 25 คนเพื่อให้ไปสอบแบบทดสอบ TOEIC แต่ปรากฏว่ามีผู้ไปสอบเพียงชุดละ 19-24 คนเท่านั้น ซึ่งเป็นจำนวนที่ค่อนข้างน้อย ดังนั้น สมการพยากรณ์คะแนนผลการทดสอบแบบทดสอบ TOEIC จึงยังมีความคลาดเคลื่อนค่อนข้างมาก

คำนิยามศัพท์ที่ใช้ในการวิจัยครั้งนี้

1. **นิสิต/นักศึกษาไทย** หมายถึงผู้สอบซึ่งเป็นคนไทยที่มีความรู้ความสามารถในระดับอุดมศึกษาและสมัครสอบแบบทดสอบวัดสมรรถภาพทางภาษาอังกฤษเพื่อวิชาชีพและการสื่อสารนานาชาติของสถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย 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 และเพื่อใช้ในวิชาชีพเป็นต้น และเนื้อหาเหล่านี้ไม่ได้กำหนดไว้ในหลักสูตรของรายวิชาใดวิชาหนึ่งโดยเฉพาะ
3. **สมรรถภาพทั่วไปทางภาษาอังกฤษเพื่อวิชาชีพและการสื่อสารนานาชาติ** หมายถึงความรู้ความสามารถของผู้สอบที่ได้จากคะแนนสอบของแบบทดสอบวัดสมรรถภาพทางภาษาอังกฤษเพื่อวิชาชีพและการสื่อสารนานาชาติของสถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย หรือ CULI-TEPOIC (Chulalongkorn University Language Institute Test of English Proficiency for Occupational and International Communication Purposes) คะแนนเหล่านี้บ่งบอกถึงความรู้ความสามารถทั่วไปทางภาษาอังกฤษของผู้สอบที่ไม่ได้ระบุไว้ในรายวิชาใดรายวิชาหนึ่งโดยเฉพาะ แต่มีความจำเป็นต่อการสื่อสารในวิชาชีพและการสื่อสารนานาชาติทั่วไป
4. **โปรแกรมคอมพิวเตอร์ที่ใช้ในการพัฒนาแบบทดสอบและการทดสอบ** หมายถึง โปรแกรมวิเคราะห์ข้อทดสอบแบบประเพณีนิยม (Classical Model) โปรแกรมวิเคราะห์ข้อทดสอบแบบราช (Rasch Model) โปรแกรมธนาคารข้อทดสอบ (Item Banking) โปรแกรมการทดสอบแบบปรับเปลี่ยน (Adaptive Testing) และโปรแกรมกำหนดระดับคะแนน (Grade Assignment)

ประโยชน์ที่คาดว่าจะได้รับ และหน่วยงานที่นำผลการวิจัยไปใช้ประโยชน์

1. นิสิต/นักศึกษาและบุคคลทั่วไปเมื่อทำการทดสอบแล้วจะได้ทราบว่าระดับสมรรถภาพทั่วไปเพื่อวิชาชีพและการสื่อสารนานาชาติทางภาษาอังกฤษของตนมีมากน้อยเพียงใด ซึ่งจะเป็นตัวกระตุ้น

- ให้นิสิต/นักศึกษาและบุคคลทั่วไปมีความสนใจ หรือแรงจูงใจในการเรียนภาษาอังกฤษมากยิ่งขึ้น และเป็นผลดีต่อสภาพการเรียนรู้การสอนภาษาอังกฤษในประเทศได้ทางหนึ่ง
2. สถาบันต่างๆ หน่วยงานของรัฐ และบริษัทห้างร้านต่าง ๆ ของเอกชนอาจใช้ผลคะแนนการทดสอบประกอบการพิจารณาเพื่อคัดเลือกรับนิสิต/นักศึกษาและบุคคลทั่วไปเข้าทำงานในหน่วยงานของตน โดยไม่ต้องเสียค่าใช้จ่ายในการพัฒนาแบบทดสอบของตนเองให้สิ้นเปลืองงบประมาณ
 3. คณะต่างๆ ของมหาวิทยาลัยอาจพิจารณาเกณฑ์ความสามารถจากระดับคะแนนสอบเป็นเกณฑ์ความสามารถขั้นต่ำของนิสิต/นักศึกษาของตนก่อนที่จะจบการศึกษา เพื่อเป็นการควบคุมคุณภาพทางภาษาของบัณฑิตของตนได้
 4. ประเทศชาติจะสามารถประหยัดเงินตราของประเทศได้ปีละจำนวนมาก หากเกณฑ์คะแนนสอบได้รับการยอมรับทั่วไป เพราะนิสิต/นักศึกษาและบุคคลทั่วไปไม่จำเป็นต้องไปทดสอบแบบทดสอบอื่นของต่างประเทศที่ต้องเสียค่าสมัครสอบแพงมาก และในขณะเดียวกันก็เป็นการส่งเสริมการใช้ภูมิปัญญาไทยในการประหยัดเงินตราของประเทศได้ด้วย
 5. สถาบันภาษาจะได้แบบทดสอบที่มีคุณภาพและได้มาตรฐานจำนวน 3 ชุด หรือ 300 ข้อ เพื่อเก็บไว้ใช้ในธนาคารข้อทดสอบสำหรับใช้ทดสอบวัดสมรรถภาพทางภาษาอังกฤษเพื่อวิชาชีพและการสื่อสารนานาชาติในโอกาสต่าง ๆ ที่ต้องการทันที
 6. สถาบันภาษาสามารถให้บริการทดสอบวัดสมรรถภาพทางภาษาอังกฤษเพื่อวิชาชีพและการสื่อสารนานาชาติแก่หน่วยงานและองค์กรต่างๆ ที่ขอรับบริการได้ทันที ทำให้งานบริการทางวิชาการด้านการทดสอบมีประสิทธิภาพและสนองความต้องการของผู้รับบริการได้มากยิ่งขึ้นนอกเหนือจากการทดสอบวัดสมรรถภาพทั่วไปทางภาษาอังกฤษอื่นที่มีอยู่แล้ว
 7. สถาบันภาษาจะได้โปรแกรมคอมพิวเตอร์จำนวน 5 โปรแกรม ซึ่งทันสมัยสำหรับใช้ในระบบการทดสอบ โดยไม่ต้องเสียเงินไปซื้อจากต่างประเทศด้วยราคาที่แพงมาก และอาจเป็นประโยชน์กับหน่วยงานหรือองค์กรอื่นที่ดำเนินการด้านการทดสอบ หรือให้บริการทดสอบได้อีกด้วย

วิธีการวิจัย

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) ในจุฬาลงกรณ์มหาวิทยาลัยและภายนอกมหาวิทยาลัย ที่สถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัยได้ประชาสัมพันธ์ให้ใช้บริการทดสอบโดยไม่ต้องเสียค่าสมัครสอบ และผู้เข้าสอบบางคนมีโอกาสดำเนินการสมัครไปสอบแบบทดสอบ 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 เพื่อเตรียมตัวสอบคัดเลือก

ไปดูงาน ณ ต่างประเทศ เพื่อเตรียมตัวสอบเลื่อนขั้นเงินเดือน หรือตำแหน่ง เพื่อให้ผ่านเกณฑ์ระดับความสามารถทางภาษาตามที่หน่วยงานกำหนดไว้ และเพื่อทดสอบความรู้ทั่วไปทางภาษาอังกฤษของตนเอง เป็นต้น การทดสอบด้วยวัตถุประสงค์ต่างๆ เหล่านี้ มีผู้เข้าสอบ จำนวนดังต่อไปนี้

ตารางที่ 1: 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

เครื่องมือการวิจัยครั้งนี้ได้แก่ 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 โดยการอิงแนวคิดและรูปแบบของแบบทดสอบ TOEIC แต่มีวัตถุประสงค์ในการทดสอบแตกต่างกัน กล่าวคือ แบบทดสอบ CULI-TEPOIC สร้างขึ้นเพื่อใช้ทดสอบสมรรถภาพทั่วไปทางภาษาอังกฤษเพื่อวิชาชีพและการสื่อสารนานาชาติ แต่แบบทดสอบ TOEIC สร้างขึ้นเพื่อใช้ทดสอบสมรรถภาพทั่วไปทางภาษาอังกฤษเพื่อการสื่อสารทั่วไป 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. ผู้วิจัยศึกษาคู่มือและขั้นตอนในการสร้างและพัฒนาแบบทดสอบให้เป็นแบบทดสอบมาตรฐานจากหนังสือบทความ 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. สร้างตารางลักษณะข้อทดสอบที่ต้องการเพื่อกำหนดสัดส่วนของข้อทดสอบ สำหรับแต่ละวัตถุประสงค์ (Create a table of specifications to determine the proportion of items on the test devoted to measuring each objective)
 2. สร้างข้อทดสอบตามลักษณะข้อทดสอบที่กำหนดไว้จำนวนหนึ่ง (Construct an initial pool of items according to the test specifications)
 3. ให้ผู้เชี่ยวชาญทางภาษาตรวจสอบความถูกต้องของภาษาและความตรงเชิงเนื้อหาของข้อทดสอบ (Use expert review and revision of items to ensure that the content and language are appropriate and that each item matches its objective)
 4. ทดลองสอบข้อทดสอบกับกลุ่มตัวอย่างที่เป็นตัวแทนของประชากรเป้าหมาย (Pilot test the items using a representative sample :เมื่อวันที่ 20 กรกฎาคม 2546)
 5. ทำการวิเคราะห์ข้อทดสอบรายข้อเพื่อตัดสินใจว่าควรปรับปรุงแก้ไขหรือตัดข้อทดสอบบางข้อทิ้ง (Conduct item analysis to inform decisions about items that need to be revised or discarded)
 6. ทดลองสอบข้อทดสอบที่ปรับปรุงแล้วอีกครั้งหนึ่ง (Pilot test the revised set of items)
 7. คำนวณหาค่าความเที่ยงและความตรงของแบบทดสอบ (Assess reliability and validity of test scores)
5. ดำเนินการสร้างและพัฒนาแบบทดสอบตามขั้นตอนต่างๆ ดังกล่าวแล้ว ยกเว้นขั้นตอนที่ 6 เนื่องจากการวิจัยครั้งนี้พบว่า มีผลวิจัยในการทดสอบแบบทดสอบแต่ละชุดจำนวนมากพอที่จะทำให้ค่าสถิติในขั้นที่ 5 คงที่เพียงพอ และค่าสถิติต่างๆ ที่ได้จากการวิเคราะห์ข้อทดสอบรายข้อ รวมทั้งแบบทดสอบทั้งฉบับสูงมากพอที่จะสามารถนำแบบทดสอบไปใช้ต่อไปได้ นอกจากนี้ยังพบว่า มีข้อทดสอบที่ควรต้องปรับปรุงแก้ไขจำนวนไม่มากนัก สามารถที่จะปรับปรุงให้ดีขึ้นแล้วนำไปใช้สำหรับการทดสอบจริงได้ในอนาคต
- ดังนั้น ในขั้นที่ 6 จึงเป็นการสุ่มผู้ที่เข้าสอบแบบทดสอบแต่ละชุดประมาณ 40 คน เพื่อไปสอบแบบทดสอบ TOEIC ในวันที่ 23 สิงหาคม 2546 เพื่อนำคะแนนผลการทดสอบมาคำนวณหาความตรงร่วมสมัย (Concurrent Validity) และสร้างตารางพยากรณ์ผลการทดสอบแบบทดสอบ TOEIC จากผลการทดสอบแบบทดสอบ CULI-TEPOIC แทนที่จะทำการทดลองสอบแบบทดสอบอีกครั้งหนึ่ง

การวิเคราะห์ข้อมูล

1. วิเคราะห์ข้อทดสอบของแบบทดสอบแต่ละชุด ด้วยวิธีประเพณีนิยม (Classical Model) เพื่อหาคุณภาพของแบบทดสอบรายข้อและทั้งฉบับ โดยใช้โปรแกรมคอมพิวเตอร์ชื่อ CTIA/Grading (สุพัฒน์ สุกมลสันต์, 2542: 30-65)
2. ทดสอบหาความเป็นเอกมิติ (Unidimensionality) ของแบบทดสอบทั้ง 3 ชุด โดยการใช้โปรแกรม 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 ทั้งนี้เพื่อใช้ประกอบการตัดสินใจในการเรียงลำดับข้อทดสอบตามความยากง่ายโดยใช้โปรแกรมคอมพิวเตอร์ชื่อ Rasch (สุพัฒน์ สุกมลสันต์, 2542: 67-84)
4. นำผลการวิเคราะห์ในขั้นที่ 1 มาศึกษาข้อดี และข้อบกพร่องที่ควรปรับปรุงแก้ไขเพื่อให้ข้อทดสอบรายข้อ

มีคุณภาพดียิ่งขึ้น โดยอาศัยเกณฑ์การคัดเลือกข้อทดสอบปรนัยแบบเลือกตอบชนิด 4 ตัวเลือกที่เป็นแบบทดสอบสมิทภาพทั่วไปดังนี้ (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. มีค่าดัชนีความตรงรายข้อ (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 และติดลบ
3. มีค่าดัชนีความตรงรายข้อ (Item Validity Index: r_{pb}) = > 0.20 และติดลบ หรือ ไม่มีนัยสำคัญ

ค. ตัวเลือกที่ควรปรับปรุง

1. มีค่าดัชนีความยาก (Difficulty Index: p) ของข้อที่ถูกต้อง < 0.15 หรือ > 0.85
2. ตัวเลือกที่ผิดมีอัตราส่วนของดัชนีความยากแตกต่างจากตัวเลือกที่ผิดอื่นมาก
3. มีค่าดัชนีอำนาจจำแนก (Discrimination Index: d) < 0.20 หรือมีค่าเป็นบวก
4. ตัวเลือกที่ผิดมีอัตราส่วนของดัชนีอำนาจจำแนกแตกต่างจากตัวเลือกที่ผิดอื่นมาก
5. มีค่าดัชนีความตรงรายข้อ (Item Validity Index: r_{pb}) < 0.25 หรือมีค่าเป็นบวก
6. ตัวเลือกที่ผิดมีอัตราส่วนของดัชนีความตรงรายข้อแตกต่างจากตัวเลือกที่ผิดอื่นมาก
7. ข้อที่ตัวเลือกที่ถูกต้องมีค่าเฉลี่ยของคะแนนเกณฑ์ (Mean Criterion Score) น้อยกว่าของตัวเลือกที่ผิด

ง. ดัชนีความเที่ยง (reliability index)

1. แบบ KR₂₀ หรือ Cronbach's Alpha = 0.70 – 0.80 หมายความว่าแบบทดสอบมีความเที่ยงในระดับดี
 2. แบบ KR₂₀ หรือ Cronbach's Alpha = 0.81 – 0.90 หมายความว่าแบบทดสอบมีความเที่ยงในระดับดีมาก
 3. แบบ KR₂₀ หรือ Cronbach's Alpha > 0.91 หมายความว่าแบบทดสอบมีความเที่ยงในระดับดีเยี่ยม
 4. แบบ KR₂₀ หรือ Cronbach's Alpha < 0.60 หมายความว่าแบบทดสอบมีความเที่ยงต่ำและควรปรับปรุงแก้ไข เว้นแต่มีข้อทดสอบจำนวนน้อย เช่น 10 เป็นต้น
5. คัดเลือกข้อทดสอบที่ดีและข้อที่ควรปรับปรุงแก้ไขจากแบบทดสอบทั้ง 3 ฉบับ เพื่อนำไปแก้ไขสำหรับใช้ในการทดสอบจริงต่อไป

สถิติที่ใช้

1. ความยากง่ายของข้อทดสอบแบบประเพณีนิยม (สุพัฒน์ สุกมลสันต์, 2538:18-19)

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

เมื่อ N_a = 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)

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

เมื่อ X_r = ค่าเฉลี่ยของคะแนนผู้สอบที่ตอบข้อทดสอบนั้นถูก
 X_w = ค่าเฉลี่ยของคะแนนผู้สอบที่ตอบข้อทดสอบนั้นผิด
 SD = ค่า SD ของคะแนนรวมของผู้สอบทั้งหมด
 p = อัตราส่วนของผู้ที่ตอบข้อทดสอบนั้นถูก
 y = ความสูงของแกน Y ในโค้งปกติ ณจุดที่ค่า P
 r_{pb} = ค่าอำนาจจำแนกแบบพอยท์ไบเซอเรียล (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_{bs}
 n = จำนวนผู้สอบทั้งหมดที่ตอบข้อทดสอบนั้น

5. ความเที่ยงของแบบทดสอบ (สุพัคณ์ 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 = อัตราส่วนของผู้ที่ตอบถูกแต่ละข้อ
 q = $(1 - p)$
 S_i^2 = ความแปรปรวนของคะแนนรวมของผู้สอบทั้งหมด
 KR_{20} = ค่าความเที่ยงของแบบทดสอบชนิดหนึ่ง

6. ความตรงร่วมสมัย (concurrent validity) ของแบบทดสอบ

ความตรงร่วมสมัยคำนวณจากค่าความสัมพันธ์ระหว่างคะแนนผลการทดสอบของแบบทดสอบ 2 ชุด ที่มุ่งทดสอบความรู้ความสามารถอย่างเดียวกัน 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. 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

สรุปผลการวิจัย

ผลของการวิจัยตามวัตถุประสงค์ทั้ง 2 ข้อมีรายละเอียดโดยสรุปดังต่อไปนี้

ส่วนที่ 1 การสร้างและพัฒนาแบบทดสอบ CULI-TEPOIC

1. ค่าสถิติโดยสรุปของแบบทดสอบ CULI-TEPOIC แต่ละชุด

ตารางที่ 2: 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} ของค่าความตรงเชิงรายข้อแบบ r_{cb}	0.383	0.459	0.441
\bar{X} ของค่าความตรงเชิงรายข้อแบบ r_{bs}	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 2 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_{cb} โดยเฉลี่ยอยู่ในระดับที่ดี (DIIA, 2003) และใกล้เคียงกันโดยเฉพาะชุดที่ 2 และ 3 ($r_{cb} = 0.292 - 0.353$) และแบบทดสอบ CULI-TEPOIC ชุดที่ 1 มีค่าความเที่ยงอยู่ในระดับดีมาก ส่วนชุดที่ 2 และ 3 มีค่าความเที่ยงอยู่ในระดับดียอดเยี่ยม(DIIA, 2003) แต่แบบทดสอบทั้ง 3 ชุดต่างก็มีค่าดังกล่าวแตกต่างกันไม่มากนัก($KR_{20} = 0.893 - 0.930$)

2. ค่าความตรงร่วมสมัย (concurrent validity) ของแบบทดสอบทั้ง 3 ชุด และสมการพยากรณ์

ตารางที่ 3: 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 3 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 3 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. ดำเนินการจัดหาและจัดจ้างผู้มีความชำนาญในการเขียนโปรแกรมคอมพิวเตอร์ด้วยภาษา Visual BASIC เพื่อใช้กับระบบ Windows และมีความเข้าใจภาษา Quick BASIC, FORTRAN และ Fox Pro เป็นอย่างดี เนื่องจากโปรแกรมต่างๆ ที่ต้องการพัฒนานั้นเดิมผู้วิจัยเขียนด้วยภาษาดังกล่าวแต่ใช้งานได้เฉพาะในระบบ 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 อนึ่งโปรแกรมทั้ง 5 โปรแกรมได้แก่

1. โปรแกรมวิเคราะห์ข้อทดสอบแบบประเพณีนิยม (Classical Model)
 2. โปรแกรมวิเคราะห์ข้อทดสอบแบบราช (Rasch Model)
 3. โปรแกรมธนาคารข้อทดสอบ (Item Banking)
 4. โปรแกรมการทดสอบแบบปรับเปลี่ยน (Adaptive Testing) และ
 5. โปรแกรมกำหนดระดับคะแนน (Grade Assignment)
3. มอปรหัสต้นฉบับ (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 และครบถ้วนเช่นเดียวกับโปรแกรมเดิม แต่ใช้งานได้สะดวกกว่าโปรแกรมชุดเดิมเนื่องจากทั้ง 5 โปรแกรม ทำการติดตั้งแยกกัน 2 ครั้งและทำงานร่วมกันได้ แทนที่จะเป็น 4 ครั้งเหมือนโปรแกรมชุดเดิมซึ่งมีบางโปรแกรมทำงานเป็นอิสระ
2. โปรแกรมที่สร้างและพัฒนาขึ้นหนึ่งโปรแกรมสามารถทำงานได้ก้าวหน้ากว่าโปรแกรมเดิม คือโปรแกรมธนาคารข้อทดสอบ กล่าวคือสามารถจัดเก็บภาพประกอบ (Graphics) ตรวจสอบรหัสความปลอดภัยของฐานข้อมูล และจัดการฐานข้อมูลตามอนุกรมวิธานได้ซับซ้อนมากกว่าโปรแกรมเดิม

ตัวอย่างต่อไปนี้แสดงผลการวิจัยของส่วนที่ 2 ได้แก่ผลการเปรียบเทียบการทำงานของโปรแกรม Classical Model กับโปรแกรม CTIA 7.0 และผลการเปรียบเทียบการทำงานของโปรแกรม Item Banking กับโปรแกรม ITB Pro

1. การเปรียบเทียบผลการทำงานของโปรแกรม Classical Model กับโปรแกรม CTIA 7.0

ตารางที่ 4: การเปรียบเทียบผลการทำงานของโปรแกรม Classical Model กับโปรแกรม CTIA 7.0 จากการวิเคราะห์แบบทดสอบชุดเดียวกัน

ค่าสถิติ *	Classical Model	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

ค่าสถิติ *	Classical Model	CTIA 7.0
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

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

ผลจากการวิเคราะห์แบบทดสอบชุดเดียวกันในตารางที่ 4 ปรากฏว่าทุกค่าในมิติต่างๆ เช่น ค่าความยากเฉลี่ย ค่าอำนาจจำแนกเฉลี่ย ค่าความตรงรายข้อเฉลี่ย และค่าความเที่ยงแบบต่างๆ มีขนาดเท่ากันพอดี และเมื่อทำการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าสถิติต่างๆ ที่ได้จากโปรแกรมทั้งสองปรากฏว่า $r_{vv} = 1.00$ ($p < 0.05$) หมายความว่าทุกค่ามีขนาดเท่ากันพอดี ซึ่งแสดงว่าโปรแกรม Classical Model ทำการคำนวณได้ค่าสถิติต่างๆ โดยสรุปจากการวิเคราะห์ข้อทดสอบรายข้อแบบประเพณีนิยมได้ถูกต้องเช่นเดียวกับโปรแกรม CTIA 7.0

2. ความสามารถในการทำงานของโปรแกรม Item Banking กับโปรแกรม ITB Pro

ตารางที่ 5: การเปรียบเทียบความสามารถในการทำงานของโปรแกรม Item Banking กับโปรแกรม ITB Pro

ความสามารถในการทำงาน	Item Banking	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. การสำรวจฐานข้อมูล	ทำได้ถูกต้อง	ทำได้
12. การตรวจรหัสความปลอดภัยของฐานข้อมูล	ทำได้ถูกต้อง	ทำไม่ได้
13. การจัดการฐานข้อมูลตามอนุกรมวิธาน	ทำได้ซับซ้อนกว่า	ทำได้
14. การใช้กับภาษาไทยและภาษาอังกฤษ	ทำได้ถูกต้อง	ทำได้
15. การจัดการต่างๆดังกล่าวกับแบบทดสอบ	ทำได้ถูกต้อง	ทำได้

จากการเปรียบเทียบความสามารถในการทำงานของโปรแกรม Item Banking กับโปรแกรม ITB Pro จากฐานข้อมูลชุดเดียวกัน ปรากฏว่าโปรแกรมใหม่ที่พัฒนาขึ้นสามารถทำงานต่างๆ ที่เกี่ยวข้องกับการทำงานของธนาคารข้อทดสอบได้ดีเช่นเดียวกับโปรแกรมเดิม เช่น การเตรียมข้อทดสอบและแบบทดสอบชนิดต่างๆ เข้าเก็บไว้ในฐานข้อมูล การปรับปรุงแก้ไข การคัดเลือกข้อทดสอบหรือแบบทดสอบ และการจัดพิมพ์สิ่งที่ได้เลือกแล้วในรูปแบบที่ต้องการได้ นอกจากนี้ยังพบว่าโปรแกรม Item Banking ยังสามารถทำงาน 3 อย่างได้ดีกว่าโปรแกรม ITB Pro คือ สามารถตรวจรหัสความปลอดภัยของฐานข้อมูลได้ และจัดเก็บภาพประกอบข้อทดสอบหรือแบบทดสอบได้ ซึ่งความสามารถดังกล่าวนี้ไม่มีในโปรแกรม ITB Pro รวมทั้งสามารถจัดการเก็บและเลือกข้อมูลจากฐานข้อมูลที่แยกประเภทตามอนุกรมวิธาน (Taxonomy) ได้ซับซ้อนมากกว่า

อภิปรายผล

ผลของการวิจัยครั้งนี้มีประเด็นที่น่าสนใจดังต่อไปนี้

1. ทำไมแบบทดสอบ CULI-TEPOIC ที่สร้างขึ้นแต่ละชุดจึงมีคุณภาพสูงและมีความตรงร่วมสมัยกับแบบทดสอบ TOEIC สูงมาก ทั้งที่มีจำนวนข้อทดสอบน้อยกว่า

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 สุกมลสันต์, 2542; Ebel, 1972) และมีความเที่ยงแบบ $KR_{20} = 0.893, 0.930$ และ 0.926 ซึ่งสูงในระดับดีมาก และดียอดเยี่ยม เนื่องจากสูงกว่าเกณฑ์ทั่วไปคือ $KR_{20} \Rightarrow 0.80$ (Jerard, 1995)

2. รูปแบบ (Format) ของแบบทดสอบ CULI-TEPOIC มีลักษณะคล้ายกับแบบทดสอบ TOEIC มาก กล่าวคือแบบทดสอบทั้งสองชนิดเป็นแบบทดสอบปรนัยชนิดเลือกตอบที่มีตัวเลือกจำนวน 4 ตัวเลือก เพื่อทดสอบความรู้และความสามารถทั่วไปทางด้านกรฟังเข้าใจความ และการเขียนและการอ่านเข้าใจความเหมือนกัน แต่ต่างกันที่จำนวนข้อทดสอบ เพราะแบบทดสอบ CULI-TEPOIC มีจำนวนข้อทดสอบส่วนละ 50 ข้อรวมเป็นข้อทดสอบทั้งหมด 100 ข้อ แต่แบบทดสอบ TOEIC มีจำนวนข้อทดสอบส่วนละ 100 ข้อรวมเป็นข้อทดสอบทั้งหมด 200 ข้อ ทำให้แบบทดสอบทั้งสองชนิดใช้เวลาสอบแตกต่างกันด้วย นอกจากนี้แบบทดสอบทั้งสองชนิดยังมุ่งทดสอบความรู้และความสามารถของผู้สอบที่คล้ายคลึงกันมาก กล่าวคือแบบทดสอบ TOEIC มีเนื้อหาที่มุ่งทดสอบความรู้และความสามารถทั่วไปในการใช้ภาษาอังกฤษเพื่อการสื่อสารในบริบทต่างๆ ในชีวิตประจำวัน แต่แบบทดสอบ CULI-TEPOIC มีเนื้อหาที่มุ่งทดสอบความรู้และความสามารถทั่วไปในการใช้ภาษาอังกฤษเพื่อการสื่อสารในบริบทต่างๆ ของวิชาชีพในชีวิตประจำวัน

3. ผู้สอบแบบทดสอบ CULI-TEPOIC แม้วามีจำนวนไม่มากนักแต่มีความรู้และความสามารถทางภาษาอังกฤษหลากหลายระดับ (Heterogeneous Group) ซึ่งอาจสังเกตได้จากพิสัย (Range) ของคะแนนสอบและดัชนีการกระจายของคะแนนสอบ (Coefficient Of Variation) กล่าวคือแบบทดสอบชุดที่ 1, 2 และ 3 มีพิสัยของคะแนน = 65, 67 และ 71 และดัชนีการกระจายของคะแนน = 22.98%, 29.19% และ

30.17% ตามลำดับ นอกจากนี้เมื่อสังเกตดูค่าความเบ้ (Skewness) ของแบบทดสอบทั้ง 3 ชุด มีค่า = -0.02, 0.09 และ 0.19 ซึ่งมีค่าใกล้เคียงกับ 0.00 แสดงว่าประมาณร้อยละ 50 ของผู้สอบแบบทดสอบแต่ละชุดมีคะแนนจากการทดสอบมากกว่าหรือน้อยกว่าคะแนนค่าเฉลี่ยของแบบทดสอบค่าต่างๆ ดังกล่าวแล้วข้างต้นแสดงให้เห็นว่าความรู้และความสามารถของผู้สอบแบบทดสอบแต่ละชุดมีความหลากหลายและอาจเป็นปัจจัยหนึ่งที่ทำให้ค่าทางสถิติของแบบทดสอบแต่ละชุดเป็นค่าที่สูงและน่าเชื่อถือ (Jerard, 1995; Cronbach, 1960: 133) เพราะว่าการกลุ่มผู้สอบที่มีความสามารถหลากหลายเป็นลักษณะของตัวแทนที่ดีของประชากรของผู้เข้ารับการทดสอบแม้ว่าจำนวนจะไม่มากนัก

4. ผู้สอบแบบทดสอบ TOEIC มีความสามารถหลากหลายเนื่องจากการได้รับการสุ่มอย่างมีระบบ (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” เพราะเหตุผลดังนี้ คือ

1. แบบทดสอบ CULI-TEPOIC และ TOEIC ต่างก็เป็นแบบทดสอบวัดสมรรถภาพทั่วไปในการใช้ภาษาอังกฤษเพื่อการสื่อสาร แต่แบบทดสอบ CULI-TEPOIC เน้นที่การสื่อสารในบริบททั่วไปในวิชาชีพ ส่วนแบบทดสอบ TOEIC เน้นที่การสื่อสารในบริบททั่วไปในชีวิตประจำวัน ดังนั้นหน้าที่ของภาษา (Language Functions) ในข้อทดสอบแม้ว่าจะมีบางส่วนเหมือนกัน แต่ก็ไม่เหมือนกันทั้งหมด
2. แบบทดสอบ CULI-TEPOIC และ TOEIC ไม่ได้เป็นแบบทดสอบคู่ขนานจริง (True Parallel Forms) กล่าวคือ ไม่ได้สร้างจากตารางกำหนดลักษณะเฉพาะ (Test Specification Table) เดียวกันจึงมีเนื้อหาต่างกัน 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 ส่วนแบบทดสอบ TOEIC มีคะแนนเฉลี่ยของผู้สอบจำนวน 65 คน = 124.318 และ S.D. = 39.28 หรือ = 62.153 และ 19.64 เมื่อเทียบเป็นคะแนนเต็มเท่ากับ 100 คะแนนเท่ากัน
3. ตารางเปรียบเทียบคะแนนเกิดจากสมการพยากรณ์ทางสถิติที่ค่าสัมประสิทธิ์สหสัมพันธ์ หรือ ค่า Beta (β) ยังไม่สมบูรณ์ กล่าวคือยังมีค่าน้อยกว่า 1.0 แม้ว่าจะมีค่าสูงมาก คือ 0.904 - 0.976 จึงยังมีความคลาดเคลื่อนในการพยากรณ์

3. ทำไมโปรแกรมที่สร้างและพัฒนาขึ้นทั้ง 5 โปรแกรมจึงสามารถทำงานได้ถูกต้องและครบถ้วนเหมือนกับโปรแกรมเดิม และมีบางโปรแกรมทำงานได้มากกว่าและดีกว่าโปรแกรมเดิม

การที่ผลการวิจัยเป็นเช่นนี้อาจเป็นเพราะเหตุผลต่อไปนี้ คือ

1. โปรแกรมใหม่ทั้ง 5 โปรแกรมสร้างจากรหัสต้นฉบับ (Source Codes) ทั้งที่เป็นสำเนาอิเล็กทรอนิกส์ (Soft Copy) และสำเนาแสดงผล (Hard Copy) ทำให้ผู้เขียนโปรแกรมที่มีความรู้ความสามารถในภาษาเดิม คือ Quick BASIC, FORTRAN และ Fox Pro และภาษาใหม่คือ Visual BASIC และ Microsoft Access 2000 สามารถปรับเปลี่ยนหรือถ่ายโอน (Transfer) ภาษาของโปรแกรมเดิม

- ให้เป็นภาษาใหม่ได้ไม่ยากนัก ดังนั้นรูปลักษณะของผลลัพธ์ (Output) ของโปรแกรมเดิมและโปรแกรมใหม่จึงเหมือนกัน
- รหัสต้นฉบับ (Source Codes) ของแต่ละโปรแกรมมีขั้นตอนวิธีการทำงาน (Algorithm) แฝงอยู่ ดังนั้นโปรแกรมใหม่ที่สร้างขึ้นจึงมีกระบวนการในการทำงานเหมือนเดิมทุกประการ ดังนั้น การทำงานของโปรแกรมใหม่ย่อมถูกต้องและครบถ้วนเหมือนกับของโปรแกรมเดิม
 - การที่มีบางโปรแกรม คือโปรแกรมธนาคารข้อทดสอบสามารถทำงานบางอย่างได้มากกว่าและดีกว่าโปรแกรมเดิมนั้น เป็นเพราะว่าภาษาคอมพิวเตอร์และเครื่องคอมพิวเตอร์ในปัจจุบันนี้มีความก้าวหน้ามากกว่าตอนที่เขียนโปรแกรมเดิมมาก กล่าวคือ ปัจจุบันนี้เครื่องคอมพิวเตอร์และภาษาที่ใช้เขียนโปรแกรมคอมพิวเตอร์บางภาษาสามารถจัดเก็บไฟล์รูปภาพ ไฟล์เสียง และไฟล์ภาพเคลื่อนไหวได้ด้วยเหตุนี้ โปรแกรมธนาคารข้อทดสอบใหม่จึงเขียนขึ้นเพื่อจัดเก็บไฟล์รูปภาพด้วยนอกเหนือจากไฟล์ตัวอักษร ส่วนไฟล์เสียง และไฟล์ภาพเคลื่อนไหวนั้นยังไม่ได้จัดเก็บเนื่องจากมีปัญหาในการสุ่มเลือกเมื่อต้องการสร้างแบบทดสอบชุดใหม่

ข้อเสนอแนะ

สืบเนื่องจากผลการวิจัยตามวัตถุประสงค์ดังกล่าวแล้วข้างต้นทั้ง 3 ข้อ ผู้วิจัยมีข้อเสนอแนะเพื่อการนำผลการวิจัยไปใช้ และเพื่อการวิจัยต่อไปในอนาคต ดังต่อไปนี้

ก. เพื่อการนำผลการวิจัยไปใช้

- เนื่องจากค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนสอบของแบบทดสอบ CULI-TEPOIC ชุดที่ 1, 2 และ 3 กับคะแนนสอบของแบบทดสอบ TOEIC ยังไม่สมบูรณ์ดังได้กล่าวมาแล้ว ทำให้การพยากรณ์คะแนนผลการทดสอบของแบบทดสอบ TOEIC จากคะแนนผลการสอบของแบบทดสอบ CULI-TEPOIC ยังมีค่าความคลาดเคลื่อน ดังนั้น ในกรณีที่ผู้สอบแบบทดสอบ CULI-TEPOIC ต้องการทราบข้อมูลเกี่ยวกับคะแนนเปรียบเทียบผลการสอบ CULI-TEPOIC และ TOEIC ของตน ผู้จัดสอบและหน่วยงานที่รับผิดชอบควรแจ้งให้ผู้สอบทราบค่าความคลาดเคลื่อนดังกล่าวด้วยทุกครั้ง รวมทั้งควรบันทึกค่าความคลาดเคลื่อนนี้ไว้เพื่อการอ้างอิงต่อไปด้วย
- 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” แต่ “สามารถเทียบเคียงกันได้”
- เนื่องจากโปรแกรมที่พัฒนาขึ้นบางโปรแกรมอาศัยแนวคิดจากทฤษฎีการตอบสนองของข้อทดสอบ (Item Response Theory) เช่น โปรแกรมการวิเคราะห์ข้อทดสอบแบบราช (Rasch Model) และโปรแกรมทดสอบแบบปรับเปลี่ยน ดังนั้น ผู้ที่จะใช้โปรแกรมดังกล่าวควรมีพื้นฐานความรู้ทางด้านการวัดและประเมินผลทางการศึกษาที่เกี่ยวข้องกับแนวคิดของทฤษฎีดังกล่าวด้วยจึงจะสามารถเข้าใจผลลัพธ์ (Output) จากการวิเคราะห์ข้อสอบจากโปรแกรมดังกล่าวได้
- หน่วยงานใดที่ให้บริการการประเมินผลสัมฤทธิ์ในการเรียน เช่น งานทะเบียนของสถานศึกษาหรือหน่วยงานที่ให้บริการการพัฒนาแบบทดสอบให้เป็นมาตรฐาน รวมทั้งให้บริการการสอบแก่บุคคลอื่น เช่น ศูนย์บริการการสอบ และสถาบันการทดสอบต่างๆ ควรจะนำโปรแกรมที่ได้จากการวิจัยครั้งนี้ไปใช้ในหน่วยงานดังกล่าว เพื่อทำให้งานประจำสามารถดำเนินไปได้อย่างถูกต้อง รวดเร็วและน่าเชื่อถือ

ได้มากยิ่งขึ้น และที่สำคัญมากก็คือ เพื่อให้งานการทดสอบ วัดและประเมินผลเป็นงานเชิงวิทยาศาสตร์ ที่สามารถตรวจสอบได้ทุกขั้นตอน

ข. เพื่อการวิจัยต่อในอนาคต

1. เนื่องจากขณะนี้แบบทดสอบ CULI-TEPOIC ทั้ง 3 ชุดอยู่ในความรับผิดชอบของศูนย์ทดสอบของสถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย ดังนั้นศูนย์ทดสอบควรดำเนินการวิจัยต่อเนื่องเพื่อพัฒนาแบบทดสอบทั้ง 3 ชุดให้มีคุณภาพดียิ่งขึ้น โดยเฉพาะการปรับแก้ข้อทดสอบจำนวนหนึ่งที่มีคุณภาพยังไม่ดี และตัวเลือกจำนวนหนึ่งที่ยังไม่ดีของข้อทดสอบที่อยู่ในเกณฑ์ที่เป็นข้อทดสอบที่ดี และถ้าเป็นไปได้ ควรทำการวิจัยเพื่อหาค่าความตรงร่วมสมัยกับแบบทดสอบมาตรฐานอีกภายหลังการปรับปรุงแก้ไขข้อทดสอบต่างๆ แล้ว
2. ศูนย์ทดสอบของสถาบันภาษาควรรับผิดชอบทำการวิจัยเพื่อสร้างและพัฒนาแบบทดสอบ CULI-TEPOIC และแบบทดสอบอื่นๆอีกตามความต้องการของสังคมและผู้ขอรับบริการ โดยก่อนที่จะให้บริการแบบทดสอบแต่ละชุดควรจะได้ผ่านการวิเคราะห์ข้อทดสอบอย่างน้อยที่สุดในระดับความตรงเชิงเนื้อหา ทั้งนี้เพื่อเป็นการประกันคุณภาพขั้นต่ำของแบบทดสอบ และหากทำได้ก็ควรทำการวิเคราะห์ข้อทดสอบเพื่อหาคุณภาพด้านอื่นๆ ที่จำเป็นต่อไป
3. ผู้สนใจควรทำการวิจัยเพื่อสร้างโปรแกรมการทดสอบแบบปรับเปลี่ยนที่ยืดเนื้อหาเป็นเกณฑ์ (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) ดังนั้นไฟล์ต่างๆเหล่านี้จึงมีความจำเป็นต่อการทดสอบในอนาคตอันใกล้

.....

บรรณานุกรม

- บึงอร สุว่างโรรส. (2527). “การศึกษาความสอดคล้องระหว่างภาษาอังกฤษธุรกิจที่สอนในมหาวิทยาลัยในประเทศไทยกับความต้องการของวงการธุรกิจ.” รายงานการวิจัย. กรุงเทพมหานคร: สถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย.
- สุพัฒน์ สุกมลสันต์ และคณะ (2529). “การประเมินผลหลักสูตรภาษาอังกฤษพื้นฐาน.” รายงานการวิจัย. กรุงเทพมหานคร: สถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย.
- สุพัฒน์ สุกมลสันต์ และคณะ. (2533). “แนวโน้มระดับสมรรถภาพทั่วไปทางภาษาอังกฤษของนิสิตจุฬาลงกรณ์มหาวิทยาลัย และบุคคลภายนอกที่สนใจปี พ.ศ. 2533” รายงานการวิจัย. กรุงเทพฯ: สถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย.
- สุพัฒน์ สุกมลสันต์. (2536). “การสอบสมรรถภาพทั่วไปทางภาษาอังกฤษครั้งที่ 3 ปี พ.ศ. 2535 ในโครงการสร้างและพัฒนาคุณภาพแบบทดสอบสมรรถภาพทั่วไปทางภาษาอังกฤษและระบบการทดสอบที่เกี่ยวข้อง.” รายงานการวิจัย. กรุงเทพมหานคร: สถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย.
- สุพัฒน์ สุกมลสันต์. (2538). การวิเคราะห์ข้อทดสอบแนวใหม่ด้วยคอมพิวเตอร์. กรุงเทพฯ: บริษัทวิทยพัฒน์ จำกัด.
- สุพัฒน์ สุกมลสันต์. (2539). ธนาคารข้อทดสอบและการทดสอบปรับเปลี่ยนด้วยคอมพิวเตอร์. กรุงเทพฯ: บริษัทวิทยพัฒน์ จำกัด.
- สุพัฒน์ สุกมลสันต์. (2542). การวิเคราะห์ข้อทดสอบและตัดเกรดด้วยคอมพิวเตอร์. กรุงเทพฯ: บริษัทวิทยพัฒน์ จำกัด.
- อัฉร วังศ์โสธร และคณะ.(2523). “การสำรวจความต้องการของสังคมต่อการใช้ภาษาอังกฤษ.” รายงานการวิจัย. กรุงเทพมหานคร: สถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย.
- อัฉร วังศ์โสธร และคณะ.(2535). “การสอบสมรรถภาพทั่วไปทางภาษาอังกฤษครั้งที่ 2 ปี พ.ศ. 2534 ในโครงการสร้างและพัฒนาคุณภาพแบบทดสอบสมรรถภาพทั่วไปทางภาษาอังกฤษและระบบการทดสอบที่เกี่ยวข้อง.” รายงานการวิจัย. กรุงเทพมหานคร: สถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย.
- Alderson, J. C. (2000). *Assessing reading*. Cambridge, UK: Cambridge University Press.
- Anderson, R. C. (1972). “How to construct achievement tests to assess comprehension.” *Review of Educational Research* 42: 145-170.
- 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.
- Bloom, B. and Others (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. New York, Toronto: Longmans, Green.
- 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.

- 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.
- 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.
- Internet World Stat. (2006). *Internet Users by Language*. Retrieved from <http://www.internetworldstats.com/stats7.htm> on August 7, 2006.
- 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.
- Lord, F.M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale: Lawrence Erlbaum.
- 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.
- Osterlind, S. J (1989). *Constructing Test Items*. Londres: Kluwer Academic Publishers.
- Runte, N. (2001). *Item analysis without complicated statistics*. Retrieved from <http://www.uleth.ca/edu/runte/tests/itemanalysis.html> on July 8, 2005.
- Valette, R. M. (1969). *Directions in foreign language testing*. NY: Modern Language Association.
- 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.
- Wright, B.D. and Mead, R.J. (1977). *BICAL: Calibrating Items and Scales with the Rasch Model (Research Memorandum 23)*. Chicago, Illinois: Statistical Laboratory, Department of Education, University of Chicago.
- 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.

