

การสร้างและพัฒนาแบบทดสอบ วัดความสามารถทางวิชาการ ของจุฬาลงกรณ์มหาวิทยาลัย สำหรับโครงการนานาชาติ

รองศาสตราจารย์ ดร. สุพัฒน์ สุขมลสันต์

ความนำ

ในปัจจุบันนี้ภาษาอังกฤษเป็นภาษาสากลที่ใช้กันอย่างแพร่หลายทั่วโลกเพื่อวัตถุประสงค์ต่างๆ กัน เช่น ประมาณร้อยละ 30 ของผู้ใช้ในระบบเครือข่ายคอมพิวเตอร์หรือประมาณ 313 ล้านคนใช้ภาษาอังกฤษเพื่อการสืบค้นข้อมูล และภาษาอังกฤษยังเป็นภาษาที่ผู้ซึ่งไม่ใช่เจ้าของภาษาใช้มากที่สุดในโลกด้วย (Internet World Stat, 2006; Vistawide, 2006) นอกจากนี้ ภาษาอังกฤษยังมีความสำคัญสำหรับการศึกษาค้นคว้าของนิสิตและนักศึกษาทั่วไปอีกด้วย

เนื่องจากจุฬาลงกรณ์มหาวิทยาลัยมีหน้าที่รับผิดชอบต่อสังคมไทยในการช่วยสร้างบัณฑิตที่มีความรู้ความสามารถที่ตอบสนองความต้องการของสังคม ซึ่งปัจจุบันนี้ต้องการบัณฑิตที่มีความรู้ความสามารถในสาขาต่างๆ ในระดับสากลมากขึ้น ดังนั้นคณะวิชาต่างๆ จึงเปิดโครงการนานาชาติขึ้นเป็นจำนวนมาก และใช้ภาษาอังกฤษเป็นสื่อในการเรียนการสอน ด้วยเหตุนี้ โครงการต่างๆ เหล่านี้จึงให้ความสำคัญต่อความรู้ความสามารถทางภาษาอังกฤษของผู้เรียนมากขึ้น และมองเห็นความจำเป็นที่จะต้อง มีแบบทดสอบที่ดีเพื่อใช้ในการสอบคัดเลือกบุคคลดังกล่าวเข้าสู่โครงการเหล่านั้น

อนึ่ง เนื่องจากในสหรัฐอเมริกา คณะกรรมการสอบคัดเลือกนักเรียนมัธยมศึกษาตอนปลายเข้าศึกษาต่อในระดับอุดมศึกษา [College Entrance Examination Board (CEEB) หรือ College Board] ใช้แบบทดสอบวัดความถนัดทางวิชาการ

รองศาสตราจารย์ ดร. สุพัฒน์ สุขมลสันต์

เป็นอาจารย์ประจำสถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย มีความชำนาญด้านการเรียนการสอนภาษาอังกฤษ การทดสอบทางภาษา สถิติเพื่อการวิจัย วิธีวิทยาการวิจัยทางภาษาศาสตร์และการศึกษา การเขียนโปรแกรมคอมพิวเตอร์ และการใช้คอมพิวเตอร์เพื่อการวิจัยและเพื่อช่วยการเรียนการสอน มีผลงานทางวิชาการหลายเล่ม และเป็นกรรมการผู้ทรงคุณวุฒิของสถาบันการศึกษาหลายแห่ง

(Scholastic Aptitude Test) หรือ SAT ในการสอบคัดเลือก ส่วนในประเทศไทย โครงการนานาชาติหลายโครงการกำหนดให้นักเรียนมัธยมศึกษาปีที่ 6 ที่มุ่งหวังเข้าเรียนในโครงการต้องสอบแบบทดสอบวัดความถนัดทางวิชาการเช่น SAT เหมือนกัน ดังนั้น เพื่อช่วยนักเรียนลดค่าใช้จ่ายในการสอบแบบทดสอบดังกล่าวซึ่งมักมีราคาสูง ผู้วิจัยจึงประสงค์ที่จะสร้างแบบทดสอบชื่อ CU-AAT (Chulalongkorn University Academic Aptitude Test for Internal Programs) ขึ้น เพื่อให้เป็นแบบทดสอบที่มีความตรงร่วมสมัย (Concurrent Validity) เทียบเท่ากับแบบทดสอบ SAT

อย่างไรก็ตาม หลังจากได้ศึกษาวรรณกรรมที่เกี่ยวข้องกับพัฒนาการของแบบทดสอบ SAT จึงทราบว่าภายหลังจาก ค.ศ. 2005 หรือปี พ.ศ.2548 แนวคิดเดิมของแบบทดสอบ SAT เปลี่ยนไปเพราะได้รับการวิพากษ์วิจารณ์ว่ามีอคติต่อผู้สอบที่มีสถานะทางสังคมและเศรษฐกิจ เพศ และเชื้อชาติแตกต่างกัน รวมทั้งข้อวิจารณ์ที่ว่าแบบทดสอบ SAT มักวัดเชาวน์ปัญญา จึงทำให้ผู้สร้างแบบทดสอบเปลี่ยนแปลงเนื้อหาของแบบทดสอบเพื่อทดสอบความสามารถด้านการใช้เหตุผลที่พัฒนาแล้ว (Developed reasoning abilities) แทนที่จะเป็นด้านความถนัดทางวิชาการ และเรียกชื่อแบบทดสอบแนวใหม่นี้ว่า Scholastic Assessment Test (แบบทดสอบวัดความสามารถทางวิชาการ) แต่ยังคงเรียกชื่อย่อแบบทดสอบนี้เหมือนเดิมคือ The SAT หรือ SAT I : Reasoning Test (Achievers Point, 2008) ดังนั้น เพื่อให้แบบทดสอบชุดใหม่ที่ผู้วิจัยสร้างขึ้นมีเนื้อหาสอดคล้องกับแบบทดสอบดังกล่าวนี้ ผู้วิจัยจึงได้ดำเนินการสร้างแบบทดสอบ Chulalongkorn University Academic Assessment Test for International Programs (CU-AAT) หรือ “แบบทดสอบวัดความสามารถทางวิชาการของจุฬาลงกรณ์มหาวิทยาลัยสำหรับโครงการนานาชาติ” แทนชื่อเดิม และแนวคิดเดิม โดยมุ่งทดสอบความสามารถด้านการใช้เหตุผลที่พัฒนาแล้ว ที่จำเป็นต้องใช้ในการศึกษาในระดับอุดมศึกษา แทนที่จะมุ่งทดสอบความถนัดทางวิชาการตามแนวคิดเดิม

วัตถุประสงค์ของการวิจัย

1. เพื่อสร้างและพัฒนาแบบทดสอบวัดความสามารถทางวิชาการสำหรับโครงการนานาชาติ (CU-AAT) ให้มีคุณภาพและได้มาตรฐานจำนวน 3 ชุด
2. เพื่อสร้างตารางเทียบคะแนนระหว่างแบบทดสอบดังกล่าวกับคะแนนแบบทดสอบ SAT ของสหรัฐอเมริกา

ขอบข่ายของการวิจัย

1. สร้างและพัฒนาแบบทดสอบ CU-AAT จำนวน 3 ชุด ให้มีเนื้อหาและรูปแบบในการวัดเช่นเดียวกับแบบทดสอบ SAT ที่จัดสร้างโดย ETS (Educational Testing Service) ของสหรัฐอเมริกา
2. แบบทดสอบ CU-AAT แต่ละชุดมีข้อทดสอบ 116 ข้อ ซึ่งรายละเอียดดังกล่าวอยู่ในหัวข้อเครื่องมือเพื่อการวิจัย
3. รูปแบบการเทียบคะแนนใช้กลุ่มตัวอย่างที่มีความสามารถไม่เท่าเทียมกันและมีข้อทดสอบร่วม (Common-Item Nonequivalent Groups Design)
4. แบบทดสอบ SAT ที่จะใช้เป็นเกณฑ์ในการเทียบคะแนนคือแบบทดสอบ SAT แนวใหม่ที่ใช้มาตั้งแต่ปี พ.ศ. 2548 ซึ่งไม่ได้มุ่งทดสอบเรื่องความถนัดทางวิชาการ และ CU-AAT ที่จะสร้างขึ้นก็ได้มุ่งทดสอบเรื่องความถนัดทางวิชาการเช่นเดียวกัน แต่ว่าแบบทดสอบทั้งสองนี้มุ่งทดสอบความสามารถทางด้านการใช้เหตุผลที่มีความสำคัญต่ออนาคตและนักศึกษาในระดับอุดมศึกษา

5. การศึกษาครั้งนี้ ผู้วิจัยไม่ได้นำคะแนนส่วนที่เป็นคณิตศาสตร์ และการเขียนเรียงความมาศึกษาด้วย เพราะว่าคุณคณิตศาสตร์ไม่ใช่ส่วนที่ผู้วิจัยสนใจ และการให้คะแนนข้อทดสอบเรียงความไม่เป็นปรนัย
6. การศึกษาครั้งนี้ประสงค์จะเทียบคะแนนระหว่างแบบทดสอบทั้ง 2 ในแนวนอน(Horizontal Equating) เท่านั้น
7. การเปลี่ยนมาตราวัด (Scale Transformation) ของแบบทดสอบ CU-AAT ชุดที่ 2 และ 3 ไปเป็นของมาตราวัดของแบบทดสอบ CU-AAT ชุดที่ 1 ใช้วิธีค่าเฉลี่ยของความยากและอำนาจจำแนกของแบบทดสอบ (Mean and Mean Method)
8. การเทียบคะแนนระหว่างแบบทดสอบ CU-AAT ชุดที่ 1 กับ SAT ใช้วิธีสมการถดถอย (Regression Equating Method) โดยอาศัยสมการทำนายคะแนนจากแบบทดสอบ CU-AAT ไปยังคะแนน SAT
9. การเทียบคะแนนระหว่างแบบทดสอบ CU-AAT ชุดที่ 2 และ 3 กับชุดที่ 1 ใช้วิธี ICC (Item Characteristic Curve Method) แบบใช้ 3-Parameter และใช้คะแนนจริง (True Score Equating Technique)
10. กลุ่มผู้สอบแบบทดสอบ CU-AAT ได้แก่ นักเรียนที่กำลังศึกษาอยู่ชั้นมัธยมศึกษาปีที่ 6 หรือเพิ่งสำเร็จการศึกษาในระดับมัธยมศึกษาตอนปลาย

ข้อตกลงเบื้องต้น

1. ผู้สอบที่เป็นนักเรียนทุกคนมีความตั้งใจในการสอบแบบทดสอบ CU-AAT เนื่องจากสมัครมาสอบตามคำประกาศเชิญชวนโดยไม่ต้องเสียค่าใช้จ่ายใดๆ ทั้งสิ้น และต่างก็ทราบว่ามีโอกาสที่จะได้รับการคัดเลือกให้ไปสอบแบบทดสอบ SAT ฉบับจริงภายหลังการสอบครั้งนี้ด้วย
2. แบบทดสอบ CU-AAT ทั้ง 3 ชุด เป็นแบบทดสอบคู่ขนานเชิงเนื้อหา เพราะสร้างจากตารางกำหนดรายละเอียดของข้อทดสอบเดียวกัน และโดยผู้สร้างคนเดียวกัน
3. เนื่องจากกลุ่มผู้สอบแบบทดสอบ CU-AAT ชุดที่ 1, 2 และ 3 สมัครใจมาสอบ และไม่ได้ทดสอบความรู้ความสามารถทางภาษาอังกฤษก่อนสอบ CU-AAT จึงอาจถือได้ว่ากลุ่มผู้สอบทั้ง 3 กลุ่มมีความรู้ความสามารถทางภาษาอังกฤษในระดับที่แตกต่างกัน และถือเป็นกลุ่มตัวอย่างแบบสุ่มอย่างง่าย

ข้อจำกัดของการวิจัยครั้งนี้

1. แบบทดสอบ CU-AAT ที่สร้างขึ้นแต่ละชุดไม่สมบูรณ์ในตัวเอง เพราะว่าคุณค่าส่วนที่เป็นด้านคณิตศาสตร์
2. เนื่องจากมีข้อจำกัดเรื่องเวลา และงบประมาณในการดำเนินการ ทำให้จัดสอบแบบทดสอบ CU-AAT กับกลุ่มตัวอย่างผู้สอบจำนวนมากหลายครั้งไม่ได้ ดังนั้นข้อทดสอบ (Test Item) จำนวนหนึ่งคงยังมีความบกพร่อง และค่าพารามิเตอร์อาจเป็นค่าที่ยังไม่คงที่
3. กลุ่มตัวอย่างที่มาสอบแบบทดสอบ CU-AAT แต่ละชุดมีจำนวนน้อยกว่า 1,000 คน และยังไม่เป็นตัวแทนที่ดีของประชากรทั้งหมด จึงอาจทำให้ค่าพารามิเตอร์ของข้อทดสอบมีความคลาดเคลื่อนอยู่บ้าง

4. ปัจจุบันยังไม่มีข้อสรุปที่แน่นอนและเป็นที่ยอมรับกันทั่วไปเกี่ยวกับขนาดของกลุ่มตัวอย่างผู้สอบความยาวของข้อทดสอบรวม และลักษณะของข้อทดสอบรวม ปัจจัยเหล่านี้อาจมีผลต่อความถูกต้องในการเทียบคะแนนระหว่างแบบทดสอบ

นิยามศัพท์เชิงปฏิบัติการสำหรับการวิจัยครั้งนี้

1. กลุ่มตัวอย่างผู้สอบ กลุ่มผู้สอบ (Testees) หรือพลวิจัย (Subjects) หมายถึงนักเรียนที่กำลังศึกษาอยู่ชั้นมัธยมศึกษาปีที่ 6 หรือเพิ่งสำเร็จการศึกษาชั้นมัธยมศึกษาตอนปลายที่สมัครมาสอบแบบทดสอบ CU-AAT
2. แบบทดสอบวัดสมรรถภาพ (Proficiency Test) หมายถึงแบบทดสอบที่มีเนื้อหาสาระที่สำคัญสำหรับการศึกษาต่อในระดับที่สูงขึ้น และเนื้อหาเหล่านี้ไม่ได้กำหนดไว้ในหลักสูตรใดหลักสูตรหนึ่งโดยเฉพาะ และแบบทดสอบที่สร้างขึ้นมุ่งเน้นเนื้อหาด้านการใช้เหตุผลที่พัฒนาแล้วโดยอาศัยภาษาอังกฤษเป็นสื่อ
3. แบบทดสอบ SAT (Scholastic Assessment Test) หมายถึงแบบทดสอบวัดความสามารถทางวิชาการที่มุ่งทดสอบความสามารถในการใช้เหตุผลที่พัฒนาแล้ว (Developed reasoning abilities) และจัดสร้างโดย ETS (Educational Testing Service) โดยการว่าจ้างของคณะกรรมการสอบคัดเลือกนักเรียนชั้นมัธยมศึกษาตอนปลายเข้าศึกษาในระดับอุดมศึกษา (College Board) และเป็นแบบทดสอบปรนัยชนิดเลือกตอบมี 5 ตัวเลือก มีจำนวน 116 ข้อ ปัจจุบันนี้เรียกว่า SAT I : Reasoning Test ซึ่งโดยปกติประกอบด้วย 2 ส่วน คือ ส่วนภาษา (Verbal Part) และส่วนคณิตศาสตร์ (Mathematics Part) แต่การศึกษาครั้งนี้ครอบคลุมเฉพาะส่วนแรกเท่านั้น
4. แบบทดสอบ CU-AAT (Chulalongkorn University Academic Assessment Test for International Programs) หมายถึงแบบทดสอบวัดความสามารถทางวิชาการที่มุ่งทดสอบความสามารถในการใช้เหตุผลที่พัฒนาแล้วที่ผู้วิจัยได้สร้างและพัฒนาขึ้น โดยมีรูปแบบ จำนวนข้อ และความมุ่งหมายเพื่อทดสอบเนื้อหาเดียวกันกับแบบทดสอบ SAT
5. การเทียบคะแนน (Test Equating) หมายถึงกระบวนการทางสถิติในการหาความสัมพันธ์ระหว่างคะแนนสอบของแบบทดสอบต่างชุดกัน แต่มุ่งทดสอบความรู้ และความสามารถอย่างเดียวกันแล้วแปลงคะแนนให้อยู่ในหน่วยเดียวกัน ก่อนนำคะแนนของแบบทดสอบต่างชุดมาเทียบกันในแนวนอน
6. แบบทดสอบที่มีคุณภาพและได้มาตรฐาน (Qualified and Standardized Test) หมายถึงแบบทดสอบที่ประกอบด้วยข้อทดสอบที่มีค่าความตรงรายข้อในระดับที่ยอมรับได้ หรือมีนัยสำคัญ ค่าความยากอยู่ในระดับค่อนข้างยาก ค่าอำนาจจำแนกอยู่ในระดับยอมรับได้ ค่าโอกาสในการเดาถูกเท่ากับหรือน้อยกว่าค่าพารามิเตอร์ของประชากร ค่าความเที่ยงอยู่ในระดับสูง และเป็นแบบทดสอบที่เหมาะสมสำหรับผู้สอบที่มีความรู้ความสามารถในระดับสูงกว่าปกติทั่วไป โดยค่าสถิติต่างๆ ดังกล่าวนี้ได้ผ่านการทดสอบกับกลุ่มตัวอย่างผู้สอบมาแล้ว

ประโยชน์ที่คาดว่าหน่วยงานที่นำผลการวิจัยไปใช้จะได้รับ

1. สถาบันภาษาจะได้แบบทดสอบที่มีคุณภาพและได้มาตรฐานจำนวน 3 ชุด เพื่อใช้ในการสอบคัดเลือกผู้สมัครเรียนโครงการนานาชาติของคณะต่างๆ ในจุฬาลงกรณ์มหาวิทยาลัย

2. สถาบันภาษาสามารถจัดเก็บแบบทดสอบดังกล่าวไว้ในธนาคารข้อทดสอบเพื่อเป็นต้นฉบับสำหรับสร้างแบบทดสอบแก่ผู้ที่ต้องการขอใช้บริการต่อไปได้
3. สถาบันภาษาได้พัฒนาศักยภาพของอาจารย์ในการสร้างแบบทดสอบมาตรฐานทุกชั้นตอน
4. ศูนย์ทดสอบทางวิชาการแห่งจุฬาลงกรณ์มหาวิทยาลัย สามารถนำแบบทดสอบ CU-AAT ดังกล่าวมาให้บริการทดสอบแก่คณะต่างๆ ที่มีโครงการนานาชาติเพื่อรับนักเรียนชั้นมัธยมศึกษาตอนปลายเข้าศึกษาต่อ หรือให้บริการแก่นักเรียนที่ต้องการทดสอบความพร้อมของตนเองว่าจะสามารถศึกษาต่อในระดับอุดมศึกษาในโครงการนานาชาติของจุฬาลงกรณ์มหาวิทยาลัยได้หรือไม่
5. ประเทศสามารถประหยัดงบประมาณได้จำนวนหนึ่ง ทั้งนี้เพราะนักเรียนสามารถสอบแบบทดสอบคู่ขนานกับแบบทดสอบที่เป็นของต่างประเทศซึ่งมีค่าใช้จ่ายสูงมากได้ในราคาประหยัด และในขณะเดียวกันก็เป็นการส่งเสริมการใช้ภูมิปัญญาไทยในการประหยัดเงินตราของประเทศด้วย

การดำเนินการวิจัย

ก. ประชากรและกลุ่มตัวอย่าง

ประชากรของการศึกษาค้นคว้านี้ได้แก่นักเรียนที่กำลังศึกษาอยู่ชั้นมัธยมศึกษาปีที่ 6 หรือเพิ่งสำเร็จการศึกษาชั้นมัธยมศึกษาตอนปลายทั่วประเทศจำนวนประมาณ 325,645 คน (กระทรวงศึกษาธิการ, 2549)

กลุ่มตัวอย่างหรือพลวิจัย (Subjects) ได้แก่กลุ่มผู้สอบที่สมัครใจมาสอบแบบทดสอบ CU-AAT ชุดที่ 1, 2 และ 3 จำนวน 224, 253 และ 285 คนตามลำดับ เนื่องจากบุคคลเหล่านี้สมัครใจมาสอบเอง และไม่ได้มีการทดสอบความสามารถทางภาษาอังกฤษก่อนการทดสอบ CU-AAT ดังนั้น จึงเชื่อว่ากลุ่มตัวอย่างเหล่านี้มีความรู้และความสามารถทางภาษาอังกฤษแตกต่างกัน และถือเป็นกลุ่มตัวอย่างแบบสุ่มอย่างง่าย

ข. เครื่องมือเพื่อการวิจัย

เครื่องมือเพื่อการวิจัยครั้งนี้ได้แก่แบบทดสอบ CU-AAT จำนวน 3 ชุด และแบบทดสอบ SAT ฉบับที่จัดสอบโดย College Board เมื่อวันที่ 22 มิถุนายน 2551

1. **แบบทดสอบ CU-AAT แต่ละชุด** เป็นแบบทดสอบปรนัยแบบเลือกตอบชนิด 5 ตัวเลือก ประกอบด้วยส่วนสำคัญดังนี้

- ก. ส่วนที่ 1: การเขียน มีทั้งหมด 49 ข้อ ใช้เวลาสอบ 50 นาที ประกอบด้วย
 1. การหาข้อผิดพลาด (Error Identification) จำนวน 18 ข้อ
 2. การปรับปรุงประโยค (Improving Sentences) จำนวน 25 ข้อ
 3. การปรับปรุงอนุচ্ছেต (Improving Paragraphs) จำนวน 6 ข้อ
- ข. ส่วนที่ 2: การอ่าน มีทั้งหมด 67 ข้อ ใช้เวลาสอบ 70 นาที ประกอบด้วย
 1. การทำประโยคให้สมบูรณ์ (Sentence Completion) จำนวน 19 ข้อ
 2. การอ่านเชิงวิพากษ์วิจารณ์ จำนวน 20 ข้อ
 3. การอ่านเชิงเปรียบเทียบ จำนวน 28 ข้อ

2. **แบบทดสอบ SAT** เป็นแบบทดสอบปรนัยแบบเลือกตอบชนิด 5 ตัวเลือก ประกอบด้วยส่วนสำคัญดังนี้

- ก. ส่วนที่ 1: การเขียน มีทั้งหมด 49 ข้อและ 1 เรื่อง ใช้เวลาสอบ 60 นาที ประกอบด้วย
1. การหาข้อผิดพลาด (Error Identification) จำนวน 18 ข้อ
 2. การปรับปรุงประโยค (Improving Sentences) จำนวน 25 ข้อ
 3. การปรับปรุงอนุচ্ছেต (Improving Paragraphs) จำนวน 6 ข้อ
 4. การเขียนเรียงความ 1 เรื่อง
- ข. ส่วนที่ 2: คณิตศาสตร์ มีทั้งหมด 54 ข้อ ใช้เวลาสอบ 70 นาที ประกอบด้วย
1. คณิตศาสตร์ทั่วไป (General Mathematics) ที่มีตัวเลือก จำนวน 44 ข้อ
 2. คณิตศาสตร์ทั่วไปที่ผู้สอบคำนวณหาคำตอบเอง จำนวน 10 ข้อ
- ค. ส่วนที่ 3: การอ่านเชิงวิพากษ์วิจารณ์ (Critical Reading) มีทั้งหมด 67 ข้อ ใช้เวลาสอบ 70 นาที ประกอบด้วย
1. การทำประโยคให้สมบูรณ์ (Sentence Completion) จำนวน 19 ข้อ
 2. การอ่านเชิงวิพากษ์วิจารณ์ จำนวน 20 ข้อ
 3. การอ่านเชิงเปรียบเทียบ จำนวน 28 ข้อ

อนึ่ง สำหรับการวิจัยครั้งนี้ ผู้วิจัยไม่ได้นำคะแนนส่วนที่เป็นคณิตศาสตร์ และการเขียนเรียงความมาศึกษาด้วย เพราะว่าคุณคณิตศาสตร์ไม่ใช่ส่วนที่ผู้วิจัยสนใจ และการให้คะแนนข้อทดสอบเรียงความไม่เป็นปรนัย

การสร้างและพัฒนาแบบทดสอบ

1. ศึกษาส่วนประกอบ แนวคิด และพัฒนาการของแบบทดสอบ SAT โดยเฉพาะฉบับที่มีการเปลี่ยนแปลงล่าสุดในปี 2548
2. ศึกษาวิธีการสร้างแบบทดสอบปรนัยชนิดเลือกตอบจากตำราและอินเทอร์เน็ต เช่น Bachman (1992), Bachman and Palmer (1990), Pike (1979) และ Alderson (2000) เป็นต้น
3. ศึกษาข้อวิพากษ์วิจารณ์ต่างๆ เกี่ยวกับแบบทดสอบ SAT จากบทความต่างๆ ทางอินเทอร์เน็ต เช่น เรื่องความมีอคติต่อสังคมและเศรษฐกิจ เพศ และเชื้อชาติ เป็นต้น
4. ศึกษาการวิเคราะห์ข้อทดสอบรายข้อตามแบบประเพณีนิยม (Classical Models) แบบการตอบสนองต่อข้อทดสอบ (Item Response Theory Models) และวิธีการเทียบคะแนนแบบต่างๆ จากตำราที่เกี่ยวข้อง เช่น Lord (1980) และ Kolen and Brennan (2004) เป็นต้น
5. ศึกษาการใช้โปรแกรมสำเร็จต่างๆ ที่ต้องใช้ในการวิจัยครั้งนี้ เช่น โปรแกรม CTIA และ Rasch (สุพัฒน์ สุขมลสันต์, 2542) และโปรแกรม IRTEQ (Han, 2007) รวมทั้งโปรแกรม SPSS Version 17.0 (SPSS, 2008) เป็นต้น

6. จัดสร้างตารางลักษณะเฉพาะของข้อทดสอบ (Test Item Specification Table) ของแบบทดสอบ CU-AAT ชุดที่ 1 จากการวิเคราะห์เนื้อหาของแบบทดสอบ SAT จากอินเทอร์เน็ต และตำราเรียนต่างๆ ที่เกี่ยวข้อง
7. สร้างแบบทดสอบ CU-AAT คู่ขนานเชิงเนื้อหาอีก 2 ชุดคือ ชุดที่ 2 และชุดที่ 3 จากตารางลักษณะเฉพาะของข้อทดสอบเดียวกัน
8. นำแบบทดสอบ CU-AAT ทั้ง 3 ชุดไปให้เจ้าของภาษาซึ่งเป็นอาจารย์สอนภาษาอังกฤษในประเทศไทยมานานกว่า 10 ปี ตรวจสอบแก้ภาษาให้ถูกต้อง
9. ประกาศเชิญชวนนักเรียนที่กำลังเรียนอยู่ชั้นมัธยมศึกษาปีที่ 6 หรือที่เพิ่งสำเร็จการศึกษาชั้นมัธยมศึกษาปลายมาสมัครสอบแบบทดสอบ CU-AAT โดยไม่ต้องเสียค่าใช้จ่ายใดๆ ในวันที่ 5 มีนาคม 2551 และระบุว่าผู้สอบทุกคนมีโอกาสดำเนินการสมัครเลือกไปสอบแบบทดสอบ SAT ฉบับจริงจำนวนประมาณ 50 คน โดยไม่ต้องเสียค่าใช้จ่ายใดๆ ทั้งสิ้น
10. จัดสอบแบบทดสอบ CU-AAT ทั้ง 3 ชุด ในวันที่ 5 มีนาคม 2551 ที่ตึกมหิตลดาภิเศก จุฬาลงกรณ์มหาวิทยาลัย ระหว่างเวลา 8.30-12.00 น. โดยจัดให้มีข้อทดสอบรวมจำนวน 25 ข้อในลักษณะข้อทดสอบร่วมภายใน (Internal Common Items)
11. นำกระดาษคำตอบของแบบทดสอบทั้ง 3 ชุดมาตรวจให้คะแนน และวิเคราะห์ข้อทดสอบรายข้อแบบประเพณีนิยม แบบ Rasch Model และแบบ 3-Parameter ตามทฤษฎีการตอบสนองต่อข้อทดสอบ (Item Response Theory) เพื่อประมวลผล และหาคุณภาพแบบทดสอบ
12. ใช้โปรแกรม SPSS คำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนผลการสอบแบบทดสอบ CU-AAT Form 1 กับคะแนนผลการสอบแบบทดสอบ SAT เพื่อหาค่าความตรงร่วมสมัย (Concurrent Validity) ของแบบทดสอบที่สร้างขึ้นใหม่
13. เรียงลำดับคะแนนของผู้สอบแบบทดสอบ CU-AAT ชุดที่ 1 จากมากที่สุดไปหาน้อยที่สุด และสุ่มเลือกผู้สอบจำนวน 50 คนที่มีคะแนนไม่ซ้ำกัน แล้วประกาศให้บุคคลเหล่านี้มาติดต่อเพื่อสมัครสอบแบบทดสอบ SAT ที่จัดสอบในวันที่ 12 มิถุนายน 2551 ณ สถานที่สอบที่ผู้สอบเลือกเอง
14. เก็บรวบรวมคะแนนสอบ CU-AAT และ SAT ของผู้สอบทั้ง 50 คน
15. ใช้โปรแกรม IRTEQ และ SPSS ประมวลผลและเปลี่ยนมาตรวัดค่าพารามิเตอร์ของแบบทดสอบ CU-AAT ชุดที่ 2 และ 3 ไปเป็นค่ามาตรวัดของแบบทดสอบ CU-AAT ชุดที่ 1
16. ใช้โปรแกรม SPSS เพื่อเทียบคะแนนของแบบทดสอบ CU-AAT ชุดที่ 1 กับคะแนนของแบบทดสอบ SAT โดยวิธีสมการถดถอย (Regression Equating Method)
17. ใช้โปรแกรม IRTEQ เพื่อเปลี่ยนมาตรวัดของค่าพารามิเตอร์ต่างๆ ของแบบทดสอบ CU-AAT ชุดที่ 2 และ 3 ให้เทียบเท่ากับมาตรวัดของแบบทดสอบ CU-AAT ชุดที่ 1 โดยวิธีใช้ค่าเฉลี่ยของความยากและอำนาจจำแนกของแบบทดสอบ (Mean and Mean Method)
18. เทียบคะแนนของแบบทดสอบ CU-AAT ชุดที่ 2 และ 3 กับคะแนนของแบบทดสอบชุดที่ 1 ด้วยวิธี ICC (Item Characteristic Curve Method) แบบใช้ 3-Parameter และคะแนนจริง (True Scores)
19. สร้างตารางเทียบคะแนนระหว่างแบบทดสอบ CU-AAT แต่ละชุดกับคะแนนแบบทดสอบ SAT

ผลการวิจัย

ผลการวิจัยมีประเด็นสำคัญ 2 ประเด็น คือ

1. คุณภาพที่ดีของแบบทดสอบ CU-AAT ทั้ง 3 ชุด

ก. ผลการวิเคราะห์ข้อทดสอบรายข้อ 3 วิธี

จากการวิเคราะห์ข้อทดสอบรายข้อ 3 วิธี คือ แบบประเพณีนิยม (Classical Model) แบบราช (Rasch Model) และแบบสามพารามิเตอร์ (3-Parameter Logistic Model) โดยใช้โปรแกรม CTIA (Classical Test Item Analysis) โปรแกรม Rasch และโปรแกรม Xcalibre (สุพัฒน์ สุขมลสันต์, 2542) ปรากฏผลดังต่อไปนี้

ตารางที่ 1: ลักษณะที่สำคัญโดยสรุปของแบบทดสอบ CU-AAT ทั้ง 3 ชุด

ที่	ค่าสถิติที่สำคัญ	แบบทดสอบ CU-AAT		
		ชุดที่ 1	ชุดที่ 2	ชุดที่ 3
1	ค่าความยาก (p) โดยเฉลี่ย	0.390	0.393	0.357
2	ค่าความยาก (Δ) โดยเฉลี่ย	14.168	14.726	15.106
3	ค่าอำนาจจำแนก (d) โดยเฉลี่ย	0.225	0.299	0.279
4	ค่าความตรงรายข้อ (r_{pb}) โดยเฉลี่ย	0.244	0.244	0.236
5	ค่าความเที่ยงแบบ KR ₂₀	0.834	0.846	0.823
6	ค่าความเที่ยงแบบ KR ₂₁	0.797	0.835	0.815
7	ค่าความเที่ยงแบบ Cronbach's Alpha	0.834	0.846	0.823
8	ค่าความเที่ยงแบบ Split-half	0.817	0.786	0.754
9	ค่าอำนาจจำแนก (α) โดยเฉลี่ย	0.91	0.99	0.88
10	ค่าความยาก (β) โดยเฉลี่ย	1.80	1.89	2.05
11	ค่าโอกาสในการเดาถูก (c) โดยเฉลี่ย	0.14	0.15	0.14
12	ค่าความสามารถที่แท้จริง (θ) ของผู้สอบโดยเฉลี่ย	-1.14	-1.17	-1.33
13	จำนวนข้อทดสอบที่ใช้ได้กับผู้สอบทุกระดับความสามารถ	62 (53.44%)	65 (56.03%)	71 (61.21%)
14	จำนวนข้อทดสอบที่ใช้ได้กับผู้สอบบางระดับความสามารถ	54 (46.56%)	51 (43.97%)	45 (38.79%)
15	สารสนเทศในการสอบที่คาดหวังไว้	12.366	12.160	9.606
16	สารสนเทศในการสอบโดยเฉลี่ย	13.321	14.404	12.196
17	เหมาะสมที่สุดสำหรับผู้มีความสามารถที่แท้จริง (θ)	1.0-2.0	1.0-2.0	1.0-2.0

ตารางที่ 1 แสดงว่าโดยสรุปแล้วแบบทดสอบ CU-AAT ทั้ง 3 ชุด มีลักษณะที่สำคัญดังนี้

1. มีความยากโดยเฉลี่ยในระดับค่อนข้างยาก กล่าวคือ $p = 0.357-0.393$ หรือ $\beta = 1.80 - 2.05$ ซึ่งเป็นลักษณะหนึ่งของแบบทดสอบ SAT ที่ College Board ของสหรัฐอเมริกาต้องการเพราะไม่ประสงค์ให้ผู้สอบจำนวนมากได้คะแนนเต็มเนื่องจากมีปัญหาในการจำแนกระดับความสามารถของผู้สอบแบบทดสอบดังกล่าว (College Board, 2005)
2. มีอำนาจจำแนกรายข้อโดยเฉลี่ยในระดับยอมรับได้ กล่าวคือ $d = 0.225-0.299$ หรือ $\alpha = 0.88 - 0.99$
3. มีโอกาสในการเดาถูกในระดับดี กล่าวคือ $c = 0.14-0.15$ ซึ่งต่ำกว่าหรือเท่ากับที่ควรจะเป็นของประชากรคือ $c = 0.15$
4. มีข้อทดสอบ (Test Item) ที่ดีที่สามารถใช้ได้กับผู้สอบทุกระดับความสามารถจำนวนมากกว่าร้อยละ 50 ของข้อทดสอบทั้งหมด คือประมาณร้อยละ 53.44-61.21 และมีข้อทดสอบที่สามารถใช้ได้ดีกับผู้สอบบางระดับความสามารถ จำนวนประมาณร้อยละ 40 กล่าวคือระหว่าง 38.79-46.56 แสดงว่าโดยเฉลี่ยแล้วแบบทดสอบ CU-AAT เป็นแบบทดสอบที่ข้อทดสอบจำนวนมากอยู่ในระดับดีมาก และข้อทดสอบอีกประมาณร้อยละ 40 เป็นข้อทดสอบที่ดี
5. มีความตรงรายข้อโดยเฉลี่ยในระดับดี คือ $r_{pb} = 0.236-0.244$
6. มีความเที่ยงแบบต่างๆ อยู่ระหว่าง 0.754-0.846 ซึ่งอยู่ในระดับดีและดีมาก (DIIA, 2003)
7. แบบทดสอบทั้ง 3 ชุดมีสารสนเทศในการทดสอบโดยเฉลี่ย (Average Test Information) สูงกว่าสารสนเทศในการทดสอบที่คาดหวังไว้ (Expected Test Information)
8. แบบทดสอบ CU-AAT ทั้ง 3 ชุด เหมาะสมที่สุดที่จะใช้ทดสอบผู้สอบที่มีระดับความสามารถโดยเฉลี่ยในระดับ $\theta = 1.0-2.0$ กล่าวคือในระดับที่สูงกว่าระดับปานกลางเล็กน้อย หรือระดับค่อนข้างเก่ง

ข. ความตรงร่วมสมัยของแบบทดสอบ CU-AAT กับแบบทดสอบ SAT

ผลการศึกษาความสัมพันธ์ระหว่างคะแนนของผู้สอบแบบทดสอบ CU-AAT ชุดที่ 1 กับคะแนนของผู้สอบแบบทดสอบ SAT ด้วยโปรแกรม SPSS Version 17.0 ปรากฏผลดังนี้

ตารางที่ 2: ความสัมพันธ์ระหว่างคะแนนผลการสอบแบบทดสอบ CU-AAT ชุดที่ 1 กับ SAT

แบบทดสอบ CU-AAT ชุดที่ 1	แบบทดสอบ SAT
---	0.882 *
	(n =50)

* $p < 0.05$

ผลการศึกษาในตารางที่ 2 ปรากฏว่าค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) ระหว่างคะแนนผลการสอบแบบทดสอบ CU-AAT ชุดที่ 1 กับแบบทดสอบ SAT ของกลุ่มผู้สอบจำนวน 50 คน = 0.882

และมีนัยสำคัญทางสถิติ ($p = 0.05$) แสดงว่าแบบทดสอบ CU-AAT ชุดที่ 1 มีความตรงร่วมสมัย (Concurrent Validity) กับแบบทดสอบ SAT ในระดับดีมาก (DIIA, 2003)

เนื่องจากแบบทดสอบ CU-AAT ทั้ง 3 ชุดสร้างจากตารางลักษณะเฉพาะของข้อทดสอบเดียวกันและโดยผู้สร้างคนเดียวกัน จึงถือได้ว่าเป็นแบบทดสอบคู่ขนานเชิงเนื้อหา ดังนั้น เมื่อแบบทดสอบ CU-AAT ชุดที่ 1 มีความตรงร่วมสมัยกับแบบทดสอบ SAT ในระดับสูงมาก จึงอนุมาน (infer) ได้ว่า แบบทดสอบ CU-AAT ชุดที่ 2 และ 3 น่าจะมีความตรงร่วมสมัยกับแบบทดสอบ SAT ในระดับที่สูงมากเช่นเดียวกัน

ดังนั้น โดยสรุปแล้วจากค่าสถิติต่างๆ ที่ปรากฏในตารางที่ 1 และ 2 แสดงว่าแบบทดสอบ CU-AAT ทั้ง 3 ชุดเป็นแบบทดสอบคู่ขนานเชิงเนื้อหาที่ดี และเหมาะสมสำหรับนำไปใช้ต่อไปได้

2. ตารางเทียบคะแนนแบบทดสอบ CU-AAT กับ SAT

ผลวิจัยครั้งนี้สามารถสร้างตารางเทียบคะแนนระหว่าง แบบทดสอบ CU-AAT ชุดที่ 1, 2 และ 3 กับแบบทดสอบ SAT ตามรายละเอียดที่ปรากฏในตารางที่ 3 ดังต่อไปนี้

ตารางที่ 3: ตารางเทียบคะแนนแบบทดสอบ CU-AAT ชุดที่ 1, 2 และ 3 กับ SAT

คะแนน CU-AAT ชุดที่ 1, 2, 3	คะแนน SAT	คะแนน CU-AAT ชุดที่ 1, 2, 3	คะแนน SAT	คะแนน CU-AAT ชุดที่ 1, 2, 3	คะแนน SAT	คะแนน CU-AAT ชุดที่ 1, 2, 3	คะแนน SAT
1, 1, 1	8	30, 47, 41	32	59, 81, 71	56	88, 102, 101	80
2, 2, 2	9	31, 49, 43	33	60, 82, 71	57	89, 102, 102	81
3, 3, 3	9	31, 51, 44	34	61, 83, 71	58	90, 103, 102	82
4, 4, 4	10	33, 52, 45	34	62, 84, 72	59	91, 103, 102	83
5, 5, 5	11	34, 53, 46	35	63, 84, 73	60	92, 104, 103	84
6, 6, 6	12	35, 55, 47	36	64, 85, 73	60	93, 104, 104	85
7, 7, 7	13	36, 56, 49	37	65, 86, 74	61	94, 105, 104	86
8, 8, 8	14	37, 58, 50	38	66, 87, 75	62	95, 106, 105	86
9, 9, 9	14	38, 59, 52	39	67, 88, 75	63	96, 106, 106	87
10, 10, 10	15	39, 60, 53	39	68, 88, 77	64	97, 107, 106	88
11, 11, 11	16	40, 61, 55	40	69, 88, 77	65	98, 107, 106	89
12, 12, 12	17	41, 63, 58	41	70, 89, 77	65	99, 108, 107	90
13, 13, 13	18	42, 63, 59	42	71, 90, 78	66	100, 108, 107	91
14, 14, 14	19	43, 65, 60	43	72, 91, , 79	67	101, 109, 108	91
15, 15, 15	19	44, 66, 61	44	73, 92, 80	68	102, 109, 108	92
16, 16, 16	20	45, 67, 62	44	74, 93, 82	69	103, 110, 109	93
17, 17, 17	21	46, 68, 63	45	75, 93, 82	70	104, 110, 109	94
18, 18, 18	22	47, 70, 63	46	76, 94, 83	70	105, 110, 110	95
19, 23, 19	23	48, 71, 63	47	77, 95, 84	71	106, 111, 110	96
20, 26, 21	24	49, 72, 64	48	78, 96, 87	72	107, 111, 110	96
21, 29, 24	24	50, 72, 64	49	79, 96, 88	73	108, 112, 111	97
22, 32, 26	25	51, 73, 65	50	80, 97, 89	74	109, 112, 111	98
23, 34, 28	26	52, 74, 67	50	81, 97, 90	75	110, 113, 112	99
24, 36, 30	27	53, 75, 68	51	82, 98, 91	75	111, 113, 112	100
25, 38, 32	28	54, 77, 68	52	83, 98, 91	76	112, 114, 113	101
26, 40, 34	29	55, 77, 69	53	84, 99, 91	77	113, 114, 113	101
27, 42, 36	29	56, 78, 70	54	85, 100, 94	78	114, 115, 114	102
28, 44, 37	30	57, 80, 70	55	86, 100, 95	79	115, 115, 114	103
29, 46, 39	31	58, 81, 70	55	87, 101, 100	80	116, 115, 114	104

หมายเหตุ ตัวเลขที่ซ้ำกันในตารางที่ 3 เกิดจากการปิดเศษทศนิยม

การอภิปรายผลการวิจัย

การวิจัยครั้งนี้ มีประเด็นที่น่าสนใจ 2 ประเด็น คือ

1. ทำไมแบบทดสอบ CU-AAT ทั้ง 3 ชุด จึงมีค่าความเที่ยงในระดับดีและดีมากเท่านั้น แทนที่จะอยู่ในระดับดีเยี่ยม

การที่จะกล่าวว่าแบบทดสอบควรมีค่าความเที่ยงสูงมากขนาดเท่าใดนั้น จะต้องพิจารณาปัจจัยต่างๆ ที่เกี่ยวข้องด้วย เช่น ความยาวของแบบทดสอบ จำนวนตัวเลือกในแบบทดสอบ วัตถุประสงค์ของการทดสอบ ผลกระทบจากการทดสอบ ตลอดจนขนาดและลักษณะทั่วไปของกลุ่มตัวอย่าง ทั้งนี้เพราะปัจจัยเหล่านี้มีผลต่อค่าความเที่ยงของแบบทดสอบ ยกตัวอย่างเช่น แบบทดสอบทางจิตวิทยาที่มีความยาวประมาณ 100 ข้อและใช้เพื่อตัดสินลักษณะของสภาพจิตของผู้เข้ารับการสอบจำนวนมากควรมีค่าความเที่ยงในระดับสูงมาก คือมีค่ามากกว่า 0.90 เช่น 0.95 เป็นต้น แต่สำหรับแบบทดสอบทั่วไปที่ครู-อาจารย์สร้างขึ้น (Teacher-made Test) ปกติแล้วมีค่าความเที่ยงประมาณ 0.60 ก็นับว่าใช้ได้แล้ว (Rudner and Schafer, 2001; Coladarci and Others, 2000)

สำหรับการวิจัยครั้งนี้ผู้วิจัยสามารถจัดสอบแบบทดสอบ CU-AAT ทั้ง 3 ชุดกับกลุ่มตัวอย่างได้เพียงครั้งเดียวด้วยข้อจำกัดเรื่องเวลา และงบประมาณในการดำเนินการ และพบว่าแบบทดสอบทั้ง 3 ชุดมีค่าความเที่ยงแบบ KR₂₀ และ Cronbach's Alpha ซึ่งถือว่าเป็นค่าความเที่ยงที่ถูกต้องและไม่มีอคติ (Wikipedia, 2008) ระหว่าง 0.823–0.846 จึงนับได้ว่าเป็นค่าความเที่ยงที่สูงในระดับที่ดีมาก (DIIA, 2003; OEA, 2005) แต่หากพิจารณาค่าความเที่ยงแบบต่างๆ ร่วมด้วย ผู้วิจัยพบว่ามีความระหว่าง 0.754–0.846 ซึ่งโดยสรุปแล้วอยู่ในระดับดีและดีมาก (DIIA, 2003)

อนึ่ง การที่แบบทดสอบทั้ง 3 ชุด มีค่าความเที่ยงในระดับดีมากเท่านั้น อาจเป็นเพราะปัจจัยต่างๆ ต่อไปนี้ (TU, 2007)

1. กลุ่มตัวอย่างไม่มีความสามารถหลากหลายเท่าที่ควรเพราะจากการสังเกตค่าสัมประสิทธิ์ของการกระจาย (Coefficient of Variation) ของคะแนนสอบแบบทดสอบ CU-AAT ทั้ง 3 ชุด พบว่ามีค่าระหว่าง 25%–35% เท่านั้น การที่กลุ่มตัวอย่างไม่มีความสามารถหลากหลายเท่าที่ควรอาจเป็นเพราะการสอบครั้งนี้เกิดจากการประกาศเชิญชวนนักเรียนระดับชั้นมัธยมศึกษาปีที่ 6 ที่สนใจมาสมัครสอบเองจึงเป็นไปได้ว่านักเรียนที่มาสมัครสอบไม่ได้เป็นตัวอย่างที่ดีของประชากรระดับชั้นมัธยมศึกษาปีที่ 6 กล่าวคือการกระจายของคะแนนสอบแบบทดสอบ CU-AAT แต่ละชุดไม่มีลักษณะเป็นโค้งปกติ ทั้งนี้สามารถสังเกตได้จากค่าความเบ้ (Skewness) และค่าความโค้ง (Kurtosis) มีค่า 1.75–1.80 และ 1.25–1.40 แทนที่จะมีค่าใกล้เคียงกับ 0 ซึ่งหมายถึงคะแนนมีการกระจายเป็นโค้งปกติ

2. แบบทดสอบ CU-AAT ทั้ง 3 ชุดมีค่าอำนาจจำแนก (d) โดยเฉลี่ยในระดับน่าพอใจหรือยอมรับได้เท่านั้น แทนที่จะมีค่าอำนาจจำแนกในระดับสูงมากกว่านี้ ทั้งนี้อาจจะเป็นเพราะกลุ่มตัวอย่างไม่มีความหลากหลายมากเท่าที่ควรดังที่ได้กล่าวมาแล้ว ด้วยเหตุนี้จึงทำให้ค่าความแปรปรวน (variance) ในการสอบไม่สูงมากเท่าที่ควร จึงทำให้ค่าอำนาจจำแนกของแบบทดสอบ CU-AAT แต่ละชุดยังไม่สูงมาก ด้วยเหตุนี้ แบบทดสอบจึงมีค่าความเที่ยงไม่สูงในระดับดีเยี่ยม กล่าวคือ สูงกว่า 0.90 (DIIA, 2003)

3. แบบทดสอบ CU-AAT แต่ละชุดมีค่าความยากโดยเฉลี่ยในระดับค่อนข้างยาก กล่าวคือ $p = 0.357$ – 0.393 หรือ $\beta = 1.80$ – 2.05 จึงทำให้การกระจายของคะแนนสอบแต่ละชุดไม่เป็นโค้งปกติ ซึ่งสามารถสังเกตได้จากความเบ้ และค่าความโค้งดังได้กล่าวแล้ว จึงอาจเป็นผลให้แบบทดสอบ CU-AAT ทั้ง 3 ชุดมีค่าความเที่ยง

ในระดับดีและดีมากเท่านั้น แต่หากจะพิจารณาว่าผู้วิจัยต้องการสร้างแบบทดสอบ CU-AAT ทั้ง 3 ชุดนี้ให้มีระดับความยากมากกว่าระดับปานกลางเล็กน้อยเพื่อให้สอดคล้องกับเจตนาของผู้สร้างแบบทดสอบ SAT (Wikipedia, 2008) การที่แบบทดสอบ CU-AAT มีค่าความเที่ยงระดับดีและดีมาก (ประมาณ 0.80) ก็ถือว่าคุ้มค่ากับการเสียบางอย่างดังกล่าวแล้ว

2. ทำไมแบบทดสอบ CU-AAT แต่ละชุดจึงมีค่าความตรงรายข้อโดยเฉลี่ยในระดับดีเท่านั้นแทนที่จะเป็นระดับดีมาก ที่เป็นเช่นนี้อาจเป็นเพราะ

1. ความสัมพันธ์ระหว่างผลการตรวจแบบทดสอบแต่ละข้อกับคะแนนรวมของแบบทดสอบแต่ละชุดโดยธรรมชาติแล้วไม่สูงมากเพราะว่าความสัมพันธ์ดังกล่าวมีลักษณะของส่วนย่อย (คือผลการตอบถูกหรือผิดของข้อทดสอบแต่ละข้อ) กับส่วนรวมใหญ่ (คือคะแนนรวมของผลทดสอบมีจำนวนมากข้อ) ค่าความสัมพันธ์มักจะต่ำ ดังนั้น เกณฑ์ทั่วไปในการพิจารณาค่าความตรงรายข้อ (Item Validity) ของแบบทดสอบคือ 0.30 ซึ่งอยู่ในระดับดี (OEA, 2005) หรือน้อยกว่าในกรณีที่เป็นแบบทดสอบที่เพิ่งพัฒนาขึ้น เช่น อาจจะเป็น 0.20-0.30 เป็นต้น (Lord, 1980) ดังนั้น เกณฑ์หนึ่งที่ใช้ในการคัดเลือกข้อทดสอบคือความมีนัยสำคัญของค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างผลการสอบรายข้อกับคะแนนรวม ซึ่งได้แก่ค่า r_{pb} ในกรณีที่การกระจายของคะแนนรวมไม่เป็นโค้งปกติ (เช่นในกรณีการศึกษาครั้งนี้) หรือค่า r_{bis} ในกรณีที่การกระจายของคะแนนรวมของผู้สอบทั้งหมดเป็นปกติ จากการศึกษาผลวิเคราะห์ข้อทดสอบของแบบทดสอบ CU-AAT ชุดที่ 1, 2 และ 3 เฉพาะตัวเลือกที่เป็นคำตอบที่ถูกต้องพบว่าค่าดังกล่าวโดยมากจะมีนัยสำคัญที่ระดับ $p = 0.05$ (เมื่อ $n > 120$ และ $t\text{-value} \geq 1.96$) ดังนั้น แสดงว่าโดยเฉลี่ยแล้วค่าความตรงของแบบทดสอบ CU-AAT ทั้ง 3 ชุดสูงในระดับดี และมีนัยสำคัญที่ระดับ 0.05 ด้วยซึ่งหมายความว่า ข้อทดสอบแต่ละข้อโดยเฉลี่ยแล้วสามารถวัดสิ่งที่ต้องการได้อย่างมีนัยสำคัญ

2. เนื่องจากแบบทดสอบ CU-AAT แต่ละชุดโดยเฉลี่ยเป็นข้อทดสอบที่ค่อนข้างยากสำหรับกลุ่มตัวอย่างตามวัตถุประสงค์ของผู้สร้าง (Wikipedia, 2008) ดังนั้นจึงอาจทำให้ความแปรปรวนของคะแนนรายข้อมีไม่มากเท่าที่ควรเมื่อเทียบกับผลการสอบของกลุ่มผู้สอบที่เป็นตัวแทนที่ดีของประชากร ดังนั้น จึงอาจทำให้ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างผลการสอบรายข้อกับคะแนนรวมต่ำกว่าที่ควร ดังที่ปรากฏในการศึกษาครั้งนี้

ข้อเสนอแนะจากการวิจัยครั้งนี้

ก. ข้อเสนอแนะเพื่อนำแบบทดสอบไปใช้

1. ผู้ที่จะนำแบบทดสอบ CU-AAT ไปใช้ควรใช้เพื่อทำการทดสอบคัดเลือก (Screening Test) กับกลุ่มผู้สอบในระดับมัธยมศึกษาตอนปลายโดยเฉพาะนักเรียนชั้นมัธยมศึกษาปีที่ 6 หรือผู้ที่เพิ่งสำเร็จการศึกษาระดับมัธยมศึกษาตอนปลายที่มีความรู้ทางภาษาอังกฤษหลากหลายและเป็นกลุ่มที่มีขนาดใหญ่ เช่น มีขนาดประมาณ 1,000 คน ทั้งนี้เพื่อให้กลุ่มตัวอย่างมีโอกาสที่จะเป็นตัวแทนที่ดีของประชากร และเพื่อจะได้ทำให้ค่าสถิติต่างๆ มีความคงที่และถูกต้องมากยิ่งขึ้น
2. ในกรณีที่สามารถดำเนินการได้ดังข้อเสนอแนะข้อที่ 1 ข้างต้น ผู้นำแบบทดสอบไปใช้ควรจะวิเคราะห์แบบทดสอบรายข้ออีกครั้งทั้งแบบประเพณีนิยม แบบ Rasch และแบบ 3-Parameter แล้วนำผลการวิเคราะห์มาเปรียบเทียบกับผลที่ปรากฏในรายงานวิจัยครั้งนี้ว่ามีการเปลี่ยนแปลงไปมากน้อยเพียงใด และควรยึดผลการวิเคราะห์ที่มีกลุ่มตัวอย่างขนาดใหญ่กว่า และเป็นตัวแทนที่ดีของประชากรเป็นเกณฑ์
3. ศูนย์ทดสอบและประเมินผลของสถาบันภาษา หรือศูนย์ทดสอบทางวิชาการแห่งจุฬาลงกรณ์มหาวิทยาลัย ที่นำแบบทดสอบที่พัฒนาขึ้นนี้ไปใช้ ควรทำการวิเคราะห์ข้อทดสอบทุกครั้ง เพื่อ

ปรับปรุงแก้ไขข้อทดสอบ และคุณภาพของแบบทดสอบให้มีมาตรฐานสูงขึ้นเสมอ

4. หากหน่วยงานที่นำแบบทดสอบ CU-AAT ไปใช้มีงบประมาณมากพอ ควรจะศึกษาเรื่องการเทียบคะแนนสอบระหว่างแบบทดสอบนี้กับแบบทดสอบ SAT อีกครั้งหนึ่ง โดยเฉพาะอย่างยิ่งเมื่อใช้แบบทดสอบนี้ไประยะเวลาหนึ่ง เช่น 3-5 ปี เพราะว่าแบบทดสอบ SAT ที่ใช้เป็นเกณฑ์อาจมีการปรับปรุงเปลี่ยนแปลงเนื้อหา และวิธีการทดสอบอีกก็ได้ เนื่องจากแบบทดสอบ SAT มีการเปลี่ยนแปลงเนื้อหาในการสอบอยู่เสมอ
5. การใช้ตารางเทียบคะแนนจากผลการวิจัยครั้งนี้ ผู้ใช้หรือผู้ให้บริการเทียบคะแนนควรคำนึงไว้เสมอว่าการเทียบคะแนนทุกวิธีมักมีความคลาดเคลื่อนในการประมาณค่า (Standard Error of Estimate) ซึ่งอาจมีค่ามากกว่าหรือน้อยกว่าค่าที่ปรากฏก็เป็นได้

ข. ข้อเสนอแนะเพื่อการวิจัยครั้งต่อไป

1. ควรทำการศึกษาเปรียบเทียบการเทียบคะแนนวิธีอื่นๆ เช่น วิธีตามแนวคิดของทฤษฎีการทดสอบแบบประเพณีนิยมแบบต่างๆ กับวิธีตามแนวคิดของทฤษฎีการตอบสนองต่อข้อทดสอบแบบต่างๆ อีกครั้ง รวมทั้งการกำหนดเงื่อนไขอื่นๆ ที่เกี่ยวข้อง เช่น ความยาวและความยากของข้อทดสอบร่วม ลักษณะของข้อทดสอบร่วม และขนาดกลุ่มตัวอย่างที่แตกต่างกัน เป็นต้น ทั้งนี้เพื่อค้นหาว่าวิธีการเทียบคะแนนลักษณะใดจึงจะทำให้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่ามีน้อยที่สุด ซึ่งจะเป็นผลดีต่อการเทียบคะแนนต่อไปในอนาคต รวมทั้งเป็นการพัฒนาศาสตร์ด้านการเทียบคะแนนแบบทดสอบต่างชุดที่มีเงื่อนไขแตกต่างกันด้วย
2. เนื่องจากปัจจุบันนี้มีโปรแกรมที่เกี่ยวข้องกับการวิเคราะห์ข้อทดสอบรายข้อ และการเทียบคะแนนจำนวนมาก ดังนั้นควรจะมีการศึกษาเปรียบเทียบผลของการใช้โปรแกรมเหล่านี้ว่า มีข้อดีและข้อเสียอย่างไร หรือไม่ ทั้งนี้เพื่อประโยชน์ของผู้ที่จะใช้โปรแกรมเหล่านี้ต่อไปในอนาคต

.....

บรรณานุกรม

- กระทรวงศึกษาธิการ. (2549). **สถิติการศึกษา**. Retrieved from http://www.MOE.org.th/edu_stat49/edu02.pdf/ on December 15, 2007.
- สุพัฒน์ สุขมลสันต์. (2542) **การวิเคราะห์ข้อทดสอบและการตัดเกรดด้วยคอมพิวเตอร์**. กรุงเทพฯ: บริษัท วิทยพัฒน์ จำกัด.
- Achievers Point. (2008). **What is SAT?** Retrieved from <http://www.achieverspoint.com/sat/sat.htm> on July 5, 2008.
- Alderson, J. C. (2000). **Assessing Reading**. Cambridge, UK: Cambridge University Press.
- Bachman, L. F. (1992). **Fundamental Considerations in Language Testing**. Oxford: Oxford University Press.
- Bachman, L. F. and Palmer, A. S. (1990). **Language Testing in Practice**. Oxford: Oxford University Press.
- Coladarci, T. and Others. (2000). **Teacher-based judgments of academic achievement: A review of literature**. *Review of Educational Research*, 3, 297-313.
- College Board. (2005). **The SAT**. Retrieved from <http://www.collegeboard.com/SAT.pdf> on June 4, 2008.
- DIIA. (2003). **Test Item Analysis and Decision Making**. Division of Instructional Innovation and Assessment, University of Texas at Austin. Retrieved from <http://www.utexas.edu/academic/diia> on June 8, 2004.
- Han, K. T. (2007). **WinGen: Windows Software that Generates IRT Parameters and Item Responses**. Retrieved from <http://www.umass.edu/remf/software/wingen/modelsR.html> on October 5, 2008.
- Internet World Stat. (2006). **Internet Users by Language**. Retrieved from <http://www.internetworldstats.com/stats7.htm> on August 7, 2006.
- Kolen, M.J. and Brennan, R.L. (2004). **Test Equating, Scaling, and Linking: Methods and Practices** (2nd Edition). Springer.
- Lord, F.M. (1980) **Applications of Item Response Theory to Practical Testing Problems**, Hillsdale, NJ.: Lawrence Erlbaum.
- OEA, (2005). **SAT: Scholastic Aptitude Test**. Retrieved from http://www.washington.edu/oea/services/testing_center/exams/sat.html on June 10, 2008.
- Pike, L. W. (1979). **An evaluation of alternative item formats for testing English as a foreign language**. *TOEFL Research Report 2*. Princeton, NJ: Educational Testing Service.
- Rudner, L.M. and Schafer, W.D. (2001). **Reliability**. ERIC Digest. Retrieved from <http://www.ericdigests.org/2002-2/reliability.htm> on July 15, 2008.
- SPSS (2008). **SPSS Version 17.0 User's Manual**

- TU. (2007). **Test and Item Analysis**. School of Medicine, Tulane University. Retrieved from http://www.som.tulane.edu/ome/helpful_hints/test_analysis.pdf on June 5, 2008.
- Vistawide. (2006). **World Languages and Cultures**. Retrieved from http://www.vistawide.com/languages/language_statistics.htm on August 7, 2006.
- Wikipedia. (2008). **SAT**. Retrieved from <http://en.wikipedia.org/wiki/SAT> on July 7, 2008.

