

บทที่ 5

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

การวิจัยเรื่อง การพัฒนาแบบสอบผลสัมฤทธิ์ทางการเรียนวิชา การวัดและประเมินผลในชั้นเรียนโดยกำหนดคะแนนมาตรฐานด้วยวิธีบูคมาร์ค การวิจัยครั้งนี้มีวัตถุประสงค์ เพื่อสร้างแบบสอบผลสัมฤทธิ์ทางการเรียนวิชา การวัดและประเมินผลในชั้นเรียน ระดับปริญญาตรี มหาวิทยาลัยบูรพา เพื่อตรวจสอบความยาก อำนาจจำแนก ความตรง และความเที่ยงของแบบสอบผลสัมฤทธิ์ทางการเรียนวิชา การวัดและประเมินผลในชั้นเรียนที่สร้างขึ้น และเพื่อหาคะแนนจุดตัดของแบบสอบที่สร้างขึ้นโดยการกำหนดมาตรฐานด้วยวิธีบูคมาร์ค

กลุ่มตัวอย่างที่ใช้ในการศึกษาแบ่งออกเป็น 2 กลุ่ม คือ แบ่งออกเป็น 2 กลุ่มคือ 1) ผู้ตัดสิน คือ อาจารย์ที่สอนรายวิชา การวัดและประเมินผลในชั้นเรียน คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา จำนวน 6 คน 2) ผู้สอบ คือ นิสิตชั้นปีที่ 3 คณะศึกษาศาสตร์ ที่ลงทะเบียนรายวิชา 400204 การวัดและประเมินผลในชั้นเรียน ปีการศึกษา 2553 โดยวิธีการสุ่มแบบยกกลุ่ม (Cluster Sampling) จำนวน 667 คน

เครื่องมือที่ใช้ในการวิจัยมีดังนี้ 1) แบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชา การวัดและประเมินผลในชั้นเรียน เป็นแบบสอบที่ผู้วิจัยได้ดำเนินการสร้างและพัฒนาขึ้นมา แบบสอบเป็นแบบเลือกตอบ 4 ตัวเลือก และแบบอัตนัย จำนวน 2 ฉบับ คือ ฉบับที่ 1 เป็นแบบสอบกลางภาค ประกอบด้วยแบบเลือกตอบจำนวน 60 ข้อ อัตนัยจำนวน 1 ข้อ ฉบับที่ 2 แบบสอบปลายภาค ประกอบด้วยแบบเลือกตอบจำนวน 50 ข้อ อัตนัยจำนวน 1 ข้อ 2) คู่มือการจัดเรียงข้อสอบ (Ordered Item Booklet: OIB) สำหรับผู้ตัดสินพิจารณาในการกำหนดคะแนนจุดตัด

การเก็บรวบรวมข้อมูล ผู้วิจัยแบ่งการเก็บรวบรวมข้อมูลออกเป็น 2 ระยะ คือ การเก็บข้อมูลระยะที่ 1 เป็นการดำเนินการเก็บข้อมูลกับผู้สอบ คือ นิสิตคณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา ที่ลงทะเบียนเรียนวิชา การวัดและประเมินผลการศึกษาในภาคเรียนที่ 1 ปีการศึกษา 2552 เพื่อพัฒนาแบบสอบวิชา การวัดและประเมินผลในชั้นเรียน โดยเป็นการทดลองใช้ครั้งที่ 1 จำนวน 5 คน เพื่อตรวจสอบความเป็นปรนัยของข้อสอบ และทดลองใช้ครั้งที่ 2 กับนิสิตจำนวน 47 คน เพื่อปรับปรุงข้อสอบ และดำเนินการเก็บข้อมูลจริงกับนิสิตคณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา ที่ลงทะเบียนเรียนวิชา การวัดและประเมินผลการศึกษาในภาคเรียนที่ 2 ปีการศึกษา 2552 จำนวน 667 คน วิธีการเก็บข้อมูลจริงคือ ดำเนินการเก็บข้อมูลโดยการจัดสอบกลางภาคเรียน และปลายภาคเรียนพร้อมกันทุกกลุ่มเรียนเพื่อพัฒนาแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชา การวัดและประเมินผลในชั้นเรียน จากนั้นผู้วิจัยนำค่าพารามิเตอร์ของข้อสอบมาจัดทำเป็นคู่มือจัดเรียงข้อสอบสำหรับผู้ตัดสินในการกำหนดมาตรฐานด้วยวิธีบูคมาร์คต่อไป การเก็บข้อมูลระยะที่ 2 เป็นการ

ดำเนินการเก็บข้อมูลกับผู้ตัดสิน ผู้ตัดสิน 6 คน ดำเนินการกำหนดมาตรฐานด้วยวิธีบูคมาร์ค โดยดำเนินการกำหนดมาตรฐานทั้งหมด 3 รอบ

การวิเคราะห์ข้อมูล ข้อมูลจากการทดลองใช้ครั้งที่ 2 วิเคราะห์ด้วยโปรแกรม EXCEL และ SPSS เพื่อหาค่าสถิติพื้นฐาน ค่าความยาก อำนาจจำแนก และความเที่ยงแบบ KR-20 ส่วนข้อมูลจากการสอบจริง วิเคราะห์ด้วยโปรแกรม Xcalibre และ Multilog เพื่อหาคุณภาพของข้อสอบ และแบบสอบ คือค่าพารามิเตอร์ของข้อสอบคือ ค่าความยาก ค่าอำนาจจำแนก และค่าการเดา และฟังก์ชันสารสนเทศของแบบสอบ และหาค่าความสามารถของผู้สอบ ที่ระดับความน่าจะเป็นที่จะตอบถูก .67 โดยโปรแกรม EXCEL ทั้งแบบสอบประเภทเลือกตอบ และอัตนัย

สรุปผลการวิจัย

1. แบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาการวัดและประเมินในชั้นเรียนประเภทเลือกตอบ **ฉบับที่ 1** มีค่าความยากอยู่ในช่วง -2.50 ถึง 3.00 คิดเป็นค่าความยากเฉลี่ยเท่ากับ 0.608 ค่าอำนาจจำแนกอยู่ในช่วง 0.49 ถึง 0.88 คิดเป็นค่าอำนาจจำแนกเฉลี่ยเท่ากับ 0.568 และค่าการเดาอยู่ในช่วง 0.11 ถึง 0.29 คิดเป็นค่าเฉลี่ยการเดาเท่ากับ 0.239 **ส่วนฉบับที่ 2** มีค่าความยากอยู่ในช่วง -1.39 ถึง 3.00 คิดเป็นค่าความยากเฉลี่ยเท่ากับ 0.358 ค่าอำนาจจำแนกอยู่ในช่วง 0.50 ถึง 0.90 คิดเป็นค่าอำนาจจำแนกเฉลี่ยเท่ากับ 0.668 และค่าการเดาอยู่ในช่วง .110 ถึง .290 คิดเป็นค่าเฉลี่ยการเดาเท่ากับ .235 จะเห็นว่า แบบสอบทั้ง 2 ฉบับนั้น ข้อสอบส่วนใหญ่ค่อนข้างยาก และสามารถจำแนกผู้สอบที่มีความสามารถแตกต่างกันได้ดี

ค่าความยาก (δ_{ij}) ค่าอำนาจจำแนก (α) ของแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาการวัดและประเมินในชั้นเรียนประเภทอัตนัย **ฉบับที่ 1** กลางภาคเรียนจำนวน 1 ข้อ **ฉบับที่ 2** ปลายภาคเรียนจำนวน 1 ข้อ พบว่า ข้อสอบทั้งสองข้อสามารถจำแนกผู้สอบที่มีความสามารถแตกต่างกันได้ดี ($\alpha = 0.98, 0.47$) ส่วนค่าความยากทั้งสองข้อ แต่ละระดับ มีค่า $\delta_{01} < \delta_{12} < \delta_{23} < \delta_{34} < \delta_{45}$ ดังนั้นค่าความยากสามารถใช้ได้ (ข้อที่ 1 แบบสอบกลางภาคมีค่าความยากอยู่ระหว่าง -2.17 กับ 0.47 ส่วนข้อที่ 1 แบบสอบปลายภาคมีค่าความยากอยู่ระหว่าง -0.98 กับ 0.59

ค่าสารสนเทศของแบบสอบ**ฉบับที่ 1** แบบสอบกลางภาคอยู่ในช่วงประมาณ 3.9 ถึง 5.0 และมีค่าสูงสุดอยู่ที่ระดับความสามารถ (θ) ประมาณ -1.0 **ฉบับที่ 2** แบบสอบปลายภาค ค่าสารสนเทศของแบบสอบอยู่ในช่วงประมาณ 3.9 ถึง 12.0 และมีค่าสูงสุดอยู่ที่ระดับความสามารถ (θ) ประมาณ -0.5 และเมื่อรวมข้อสอบทั้ง 2 ฉบับ 100 ข้อ ค่าสารสนเทศของแบบสอบอยู่ในช่วงประมาณ 2.0 ถึง 39.0 และมีค่าสูงสุดอยู่ที่ระดับความสามารถ (θ) ประมาณ 1.5 จะเห็นว่าแบบสอบทั้ง 2 ฉบับ เหมาะสำหรับนิสิตค่อนข้างเก่ง

ความตรงตามสภาพของแบบสอบ ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างระดับผลการเรียนเฉลี่ยกับคะแนนสอบ วิชาวัดและประเมินผลในชั้นเรียนกลางภาคเท่ากับ 0.576 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างระดับผลการเรียนเฉลี่ยกับคะแนนสอบ วิชาวัดและประเมินผลในชั้นเรียนปลายภาคเท่ากับ 0.422 มีระดับนัยสำคัญที่ระดับ .01

2. คะแนนจุดตัดของแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาการวัดและประเมินในชั้นเรียน เมื่อผู้ตัดสินกำหนดคะแนนจุดตัด 7 ระดับ ด้วยวิธีบูคมาร์ค รอบที่ 1 มีคะแนนจุดตัดที่เป็นคะแนนดิบ และคะแนนความสามารถระดับดีเยี่ยม (A) เท่ากับ 109.000, 1.453 ตามลำดับ ระดับดีมาก (B+) เท่ากับ 90.500, 1.293 ตามลำดับ ระดับดี (B) เท่ากับ 79.000, 1.178 ตามลำดับ ระดับดีพอใช้ (C+) เท่ากับ 63.500, 1.128 ตามลำดับ ระดับพอใช้ (C) เท่ากับ 47.000, 1.078 ตามลำดับ ระดับอ่อน (D+) เท่ากับ 26.500, 1.008 ตามลำดับ และระดับอ่อนมาก (D) เท่ากับ 13, 0.993 ตามลำดับ

รอบที่ 2 มีคะแนนจุดตัดที่เป็นคะแนนดิบ และคะแนนความสามารถระดับดีเยี่ยม (A) เท่ากับ 109.000, 1.453 ตามลำดับ ระดับดีมาก (B+) เท่ากับ 89.000, 1.283 ตามลำดับ ระดับดี (B) เท่ากับ 78.000, 1.188 ตามลำดับ ระดับดีพอใช้ (C+) เท่ากับ 65.000, 1.133 ตามลำดับ ระดับพอใช้ (C) เท่ากับ 49.000, 1.078 ตามลำดับ ระดับอ่อน (D+) เท่ากับ 29.000, 1.013 ตามลำดับ และระดับอ่อนมาก (D) เท่ากับ 12.000, .933 ตามลำดับ

รอบที่ 3 มีคะแนนจุดตัดที่เป็นคะแนนดิบ และคะแนนความสามารถระดับดีเยี่ยม (A) เท่ากับ 109.000, 1.453 ตามลำดับ ระดับดีมาก (B+) เท่ากับ 91.000, 1.293 ตามลำดับ ระดับดี (B) เท่ากับ 83.000, 1.203 ตามลำดับ ระดับดีพอใช้ (C+) เท่ากับ 65.000, 1.133 ตามลำดับ ระดับพอใช้ (C) เท่ากับ 49.500, 1.093 ตามลำดับ ระดับอ่อน (D+) เท่ากับ 30.000, 1.0130 ตามลำดับ และระดับอ่อนมาก (D) เท่ากับ 13.000, .933 ตามลำดับ

3. ความตรงของการกำหนดคะแนนจุดตัด ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างระดับผลการเรียนของนิสิต (GPAX) กับระดับผลการตัดสินที่กำหนดด้วยวิธีการบูคมาร์ค มีความสัมพันธ์กันปานกลาง ($r = .524$) อย่างมีนัยสำคัญทางสถิติที่ระดับ .01



อภิปรายผลการวิจัย

การวิจัยครั้งนี้มีจุดมุ่งหมายคือ เพื่อสร้างแบบสอบผลสัมฤทธิ์ทางการเรียนวิชา การวัด และประเมินผลในชั้นเรียน เพื่อตรวจสอบความยาก อำนาจจำแนก ความตรง และความเที่ยงของแบบสอบผลสัมฤทธิ์ทางการเรียนที่สร้างขึ้น และเพื่อหาคะแนนจุดตัดของแบบสอบที่สร้างขึ้นโดยการกำหนดมาตรฐานด้วยวิธีบูคมาร์ค ดังนั้นผู้วิจัยจึงขออภิปรายผลดังนี้

1. คุณภาพของแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชา การวัดและประเมินในชั้นเรียน

1.1 ความยาก ข้อสอบเลือกตอบฉบับที่ 1 จำนวน 60 ข้อ มีค่าความยากอยู่ระหว่าง -2.50 ถึง 3.00 คิดเป็นค่าความยากเฉลี่ยเท่ากับ 0.608 เมื่อพิจารณาจะเห็นว่า ข้อสอบส่วนใหญ่ค่อนข้างยาก โดยจะมีข้อสอบที่ยากมาก (ค่าความยากสูงกว่า 2.50) จำนวน 8 ข้อ และง่ายมาก (ค่าความยากน้อยกว่า -2.50) จำนวน 1 ข้อ ซึ่งไม่เป็นไปตามเกณฑ์ ส่วนฉบับที่ 2 มีจำนวน 50 ข้อ มีค่าความยากอยู่ในช่วง -1.39 ถึง 3.00 คิดเป็นค่าความยากเฉลี่ยเท่ากับ 0.358 เมื่อพิจารณาจะเห็นว่า ข้อสอบส่วนใหญ่ค่อนข้างยากแต่ง่ายกว่าข้อสอบฉบับที่ 1 โดยจะมีข้อสอบที่ยากมาก (ค่าความยากสูงกว่า 2.50) จำนวน 5 ข้อ ซึ่งไม่เป็นไปตามเกณฑ์ โดยเกณฑ์ที่ใช้คัดเลือกข้อสอบที่มีคุณภาพดีคือค่าความยากอยู่ระหว่าง -2.50 ถึง +2.50 (ศิริชัย กาญจนวาสี, 2545) อาจเป็นไปได้ว่าข้อสอบดังกล่าวยังมีความกำกวม ถึงแม้จะผ่านกระบวนการพัฒนาข้อสอบมาแล้ว เนื่องจากทดลองครั้งที่ 1 กับนิสิตเพียง 5 คน

สำหรับค่าความยากของข้อสอบอัตนัยทั้งสองข้อในแบบสอบทั้งสองฉบับ (ข้อละฉบับ) แต่ละข้อมีค่าความยากประจำระดับคะแนน (มี 5 ระดับคะแนน) มีค่า $\delta_{01} < \delta_{12} < \delta_{23} < \delta_{34} < \delta_{45}$ จะเห็นว่าข้อสอบทั้งสองข้อ ระดับคะแนน 1 คะแนนมีความยากน้อยที่สุด และระดับคะแนน 5 คะแนนมีความยากมากที่สุด ซึ่งจัดว่าข้อสอบใช้ได้ อาจเนื่องมาจากผู้วิจัยมีเกณฑ์การให้คะแนนที่ชัดเจนสำหรับให้อาจารย์ประจำวิชาตรวจให้คะแนน

1.2 อำนาจจำแนก ข้อสอบเลือกตอบฉบับที่ 1 จำนวน 60 ข้อ ค่าอำนาจจำแนกอยู่ในช่วง 0.49 ถึง 0.88 คิดเป็นค่าอำนาจจำแนกเฉลี่ยเท่ากับ 0.568 ส่วนฉบับที่ 2 มีจำนวน 50 ข้อ ค่าอำนาจจำแนกอยู่ในช่วง 0.50 ถึง 0.90 คิดเป็นค่าอำนาจจำแนกเฉลี่ยเท่ากับ 0.668 จะเห็นว่าข้อสอบสามารถจำแนกนิสิตได้เกือบทุกข้อ มีเพียง 2 ข้อที่มีค่าอำนาจจำแนก 0.49 โดยเกณฑ์ที่ใช้คัดเลือกข้อสอบที่มีคุณภาพดีคือค่าอำนาจจำแนกอยู่ระหว่าง +0.50 ถึง +2.50 (ศิริชัย กาญจนวาสี, 2545) ซึ่งก็มีเพียง 2 ข้อ และค่าอำนาจจำแนกก็เกือบ 0.50 ดังนั้นข้อสอบส่วนใหญ่สามารถจำแนกนิสิตได้

1.3 ความเที่ยงของแบบสอบ ค่าความเที่ยงของแบบสอบพิจารณาจากค่าสารสนเทศของแบบสอบฉบับที่ 1 มีค่าสูงสุดอยู่ที่ระดับความสามารถ (θ) ประมาณ -1.0 ฉบับที่ 2 มีค่าสูงสุดอยู่ที่ระดับความสามารถ (θ) ประมาณ -0.5 จะเห็นว่าแบบสอบทั้ง 2 ฉบับ เหมาะสำหรับนิสิตปานกลาง เมื่อพิจารณาค่าความเที่ยงอิงเกณฑ์สูตรของลิวิงสตัน (Livingston) ของแบบสอบฉบับที่ 1 คือ 0.983 และแบบสอบฉบับที่ 2 คือ 0.984 เมื่อพิจารณาตามทีนันนาลลี และเบอร์นสไตน์ (Nunnally and Bernstein, 1994) กล่าวว่า แบบสอบที่ต้องการตัดสินใจที่สำคัญ ค่าความเที่ยงขั้นต่ำคือ .90 ส่วนค่าความเที่ยง .95 ขึ้นไป เป็นระดับที่ได้มาตรฐาน แสดงว่าแบบสอบทั้งสองฉบับมีค่าความเที่ยงที่ได้มาตรฐาน

1.4 ความตรงตามสภาพของแบบสอบ ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างระดับผลการเรียนเฉลี่ยกับคะแนนสอบ วิชาวัดและประเมินผลในชั้นเรียนกลางภาคเท่ากับ 0.576 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างระดับผลการเรียนเฉลี่ยกับคะแนนสอบ วิชาวัดและประเมินผลในชั้นเรียนปลายภาคเท่ากับ 0.422 มีระดับนัยสำคัญที่ระดับ .01 แสดงว่าแบบสอบที่สร้างขึ้นทั้งสองฉบับมีความตรงตามสภาพอยู่ในระดับปานกลาง อาจเนื่องจากระดับคะแนนเฉลี่ยเป็นการคิดหลาย ๆ วิชามารวมกันที่มีทั้งวิชาบรรยายและวิชาปฏิบัติ ส่วนวิชาการวัดและประเมินเป็นเพียงวิชาหนึ่งเท่านั้นที่เป็นวิชาบรรยาย

2. การกำหนดคะแนนจุดตัดด้วยวิธีบูคมาร์ค

2.1 ผลคะแนนจุดตัดของแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชา การวัดและประเมินผลในชั้นเรียน เมื่อผู้ตัดสินกำหนดคะแนนจุดตัดด้วยวิธีการบูคมาร์ค จำนวน 3 รอบ มีคะแนนจุดตัด 7 ระดับคือ ระดับดีเยี่ยม (A) ระดับดีมาก (B+) ระดับดี (B) ระดับดีพอใช้ (C+) ระดับพอใช้ (C) ระดับอ่อน (D+) และระดับอ่อนมาก (D) พบว่า รอบที่ 1 ส่วนเบี่ยงเบนมาตรฐานของคะแนนจุดตัดที่เป็นคะแนนดิบ มีค่าระหว่าง 3.146 ถึง 14.675 มีส่วนเบี่ยงเบนมาตรฐานเฉลี่ยเท่ากับ 9.272 รอบที่ 2 มีค่าส่วนเบี่ยงเบนมาตรฐานคะแนนจุดตัดที่เป็นคะแนนดิบระหว่าง 1.890 ถึง 7.675 มีส่วนเบี่ยงเบนมาตรฐานเฉลี่ยเท่ากับ 4.601 และรอบที่ 3 มีค่าส่วนเบี่ยงเบนมาตรฐานคะแนนจุดตัดที่เป็นคะแนนดิบระหว่าง 0.408 ถึง 5.320 มีส่วนเบี่ยงเบนมาตรฐานเฉลี่ยเท่ากับ 3.678 จะเห็นได้ว่าวิธีการกำหนดมาตรฐานด้วยวิธีบูคมาร์ค ส่วนเบี่ยงเบนมาตรฐานของคะแนนจุดตัดของผู้ตัดสินในรอบที่ 1 ต่ำกว่ารอบที่ 2 และรอบที่ 3 ส่วนรอบที่ 2 และ รอบที่ 3 มีค่าใกล้เคียงกัน แสดงให้เห็นว่า ผู้ตัดสินมีความสอดคล้องกันสูงในการกำหนดคะแนนจุดตัดด้วยวิธีบูคมาร์คเมื่อจำนวนรอบเพิ่มขึ้น ดังที่เบอร์ก (Berk, 1995) กล่าวว่า ความแปรผันระหว่างผู้ตัดสินต่ำเป็นผลทางบวก นอกจากนี้ผลการวิจัยครั้งนี้ยังสอดคล้องกับงานวิจัยของ บาคเคนดัล (Buckendahl, 2002)

และดาวเบอร์ และคณะ (Dawber et al., 2002) ที่พบว่าส่วนเบี่ยงเบนมาตรฐานของคะแนนจุดตัดมีค่าลดลงเมื่อเทียบกับรอบที่ 1 และรอบที่ 2 และส่วนเบี่ยงเบนมาตรฐานรอบที่ 2 และรอบที่ 3 มีความแตกต่างกันเล็กน้อย

จากที่กล่าวมาจะเห็นว่าคะแนนจุดตัดในรอบสุดท้ายมีส่วนเบี่ยงเบนมาตรฐานไม่มาก แสดงให้เห็นว่าผู้ตัดสินมีการกำหนดคะแนนจุดตัดแต่ละระดับใกล้เคียงกัน นั่นคือกระบวนการบู้คมาร์คจะช่วยให้ผู้ตัดสินกำหนดคะแนนจุดตัดให้มีความเที่ยงตรงภายในมากขึ้น โดยมีเครื่องมือช่วยคือ คู่มือจัดเรียงข้อสอบที่ช่วยให้ผู้ตัดสินกำหนดคะแนนจุดตัดแต่ละระดับได้ง่ายขึ้น เพราะมีข้อมูลให้ในแต่ละหน้า คือ มีข้อสอบให้เห็น ซึ่งเป็นโจทย์คำถามที่เรียงจากข้อง่ายสุดไปยากสุด ดังนั้นผู้ตัดสินสามารถเปรียบเทียบข้อสอบแต่ละข้อทีละหน้า ทำให้ง่ายต่อการตัดสินใจเพื่อกำหนดว่าหน้าไหนจะเป็นคะแนนจุดตัดในระดับนั้น นอกจากนี้ยังเปิดโอกาสให้ผู้ตัดสินปรับคะแนนจุดตัดเมื่อได้รับฟังความคิดเห็นจากผู้ตัดสินคนอื่น ๆ ได้ 2 รอบ ซึ่งเทคนิคดังกล่าวเป็นการนำเทคนิคเดลฟายมาใช้ในการกำหนดคะแนนด้วยบู้คมาร์ค ด้วยเหตุผลที่ผู้วิจัยกล่าวข้างต้น เป็นการสนับสนุนคำกล่าวของ บาคเคนดัล (Buckendahl, 2002) และดาวเบอร์ และคณะ (Dawber et al., 2002) ว่า วิธีการกำหนดมาตรฐานด้วยวิธีบู้คมาร์ค เป็นวิธีการที่นิยมใช้กันอย่างกว้างขวาง เนื่องจากวิธีการนี้ช่วยผู้ตัดสินในการคิดกำหนดคะแนนจุดตัดให้ง่ายในการทำ และมีหลักฐานสำคัญที่ช่วยผู้ตัดสินเข้าใจและเชื่อมั่นกระบวนการในการใช้

เมื่อผู้ตัดสินกำหนดคะแนนจุดตัดจากคะแนนเต็ม 120 คะแนน พบว่าคะแนนจุดตัด 7 ระดับ มีดังนี้คือ คะแนนจุดตัดระดับดีเยี่ยม (A) 109 คะแนนขึ้นไป คะแนนจุดตัดระดับดีมาก (B+) 91 คะแนนขึ้นไป คะแนนจุดตัดระดับระดับดี (B) 83 คะแนนขึ้นไป คะแนนจุดตัดระดับระดับดีพอใช้ (C+) 65 คะแนนขึ้นไป คะแนนจุดตัดระดับระดับพอใช้ (C) 49.5 คะแนนขึ้นไป คะแนนจุดตัดระดับระดับอ่อน (D+) 30 คะแนนขึ้นไป คะแนนจุดตัดระดับอ่อนมาก (D) 13 คะแนนขึ้นไป จากที่กล่าวมาเป็นที่น่าสังเกตว่าคะแนนจุดตัดบู้คมาร์คจะให้คะแนนจุดตัดระดับอ่อนมากต่ำมากคือ 13 คะแนน และคะแนนระดับดีเยี่ยมมีคะแนนจุดตัดสูงมากคือ 109 คะแนน จากผลการวิจัยดังกล่าวสอดคล้องกับการตั้งสมมติฐานของ กรีนและคณะ (Green et al., 2003); ยินและซูลส์ (Yin & Schulz, 2005) พบว่า วิธีบู้คมาร์คจะให้คะแนนจุดตัดต่ำที่สุดเมื่อเทียบกับวิธีการอื่น

2.2. เปอร์เซ็นต์การผ่านของนิสิตในแต่ละระดับมาตรฐานที่กำหนดมาตรฐานด้วยวิธีบู้คมาร์ค พบว่า มีนิสิตระดับดีเยี่ยม (A) 0.3% นิสิตระดับดีมาก (B+) 4.5% นิสิตระดับดี (B) 13.5% นิสิตระดับดีพอใช้ (C+) 41.5% นิสิตระดับพอใช้ (C) 22.8% นิสิตระดับอ่อน (D+) 17.1% นิสิตระดับอ่อนมาก (D) 0.3% และระดับตก (F) ไม่มีนักเรียนตก

เมื่อพิจารณาจะเห็นว่า เปอร์เซ็นต์การผ่านของนักเรียนที่กำหนดมาตรฐานด้วยวิธี บัญญัติ ส่วนใหญ่นักเรียนเกือบครึ่งหนึ่งคือ ประมาณร้อยละ 40 อยู่ระดับดีพอใช้ (C+) และยังมี ข้อสังเกตอีกว่า เมื่อรวมเปอร์เซ็นต์ของนิสิตที่อยู่ดีพอใช้ (C+) และนิสิตระดับพอใช้ (C) มีประมาณ 60% จากที่กล่าวมาจะเห็นว่า การกำหนดมาตรฐานด้วยวิธีบัญญัติ นักเรียนส่วนใหญ่อยู่ในระดับ มาตรฐานกลาง ๆ คือ C และ C+ เมื่อคำนวณค่าความเบ้ของข้อมูลระดับผลการเรียนเมื่อกำหนด คะแนนจุดตัดด้วยวิธีบัญญัติมีค่า -1.14 แสดงว่า เบ้ลบ นั่นคือ ระดับคะแนนส่วนใหญ่มาด้าน C +, B, B+

นอกจากนี้ผู้วิจัยยังได้สัมภาษณ์ผู้ตัดสิน พบว่า วิธีบัญญัติเป็นวิธีที่ง่ายต่อการ นำไปใช้ ขั้นตอนไม่ยุ่งยาก สะดวก ประหยัดเวลา เนื่องจากการพิจารณาข้อสอบทำได้รวดเร็วเพราะ มีการจัดเรียงข้อสอบตามความยากของข้อสอบจากง่ายไปหายาก ทำให้การพิจารณาตัดสินคะแนน จุดตัดเป็นไปอย่างรวดเร็ว ผู้ตัดสินมีความคล่องตัวในการดำเนินงานตัดสิน จากการสัมภาษณ์ ดังกล่าวผู้วิจัยคิดว่า กระบวนการที่ให้ผู้ตัดสินทำการเปิดคู่มือจัดเรียงข้อสอบที่ละหน้าพอถึงระดับที่ นักเรียนคาบเส้นระดับนั้นไม่สามารถทำได้ก็ค้นหนังสือไว้ จากนั้นก็ทำต่อไปจนครบ 7 ระดับ ซึ่งวิธีนี้ ผู้ตัดสินใช้เวลาไม่มาก ทำให้วิธีบัญญัติน่าจะเหมาะสำหรับการประเมินที่มีคะแนนจุดตัดหลาย ระดับ และมีข้อสอบจำนวนมาก และมีคู่มือจัดเรียงข้อสอบ และการประชุมกลุ่มแลกเปลี่ยนความ คิดเห็น ทำให้ช่วยจัดระบบการคิดของผู้ตัดสิน ส่งผลต่อการกำหนดคะแนนจุดตัดของผู้ตัดสิน สอดคล้องใกล้เคียงกัน หรือบางคนตรงกัน ทำให้ค่าความแปรปรวนของผู้ตัดสินต่ำ ส่งผลให้ค่า ความเที่ยงของการกำหนดมาตรฐานสูงนั่นเอง ซึ่งสอดคล้องกับคำกล่าวของ ลิวิส (Lewis et al., 1998) ว่า วิธีบัญญัติพัฒนาขึ้น จากข้อจำกัดบางประการของวิธีการกำหนดมาตรฐานวิธีอื่น ๆ โดยมีจุดเด่นคือ สามารถกำหนดคะแนนจุดตัดหลายระดับได้ ใ้กับรูปแบบข้อสอบหลายแบบเช่น แบบ เลือกตอบ และแบบอัตนัย และสุดท้ายผู้ตัดสินใช้การคิด (Cognitive Tasks) ในการกำหนดคะแนน จุดตัดได้ง่ายขึ้น

แต่ผู้วิจัยยังอีกประเด็นหนึ่งคือ ขั้นตอนการเตรียมคู่มือจัดเรียงข้อสอบใช้เวลานานจึงแล้ว เสร็จเนื่องจากต้องจัดเตรียมสารสนเทศทั้งค่าความยาก คำนวณค่าความสามารถรายข้อ จัดข้อสอบ ให้อยู่ที่ละหน้า จะเห็นว่าจุดอ่อนของวิธีบัญญัติ อยู่ที่การทำคู่มือจัดเรียงข้อสอบใช้เวลาใน การดำเนินการ และเสียค่าใช้จ่ายสูงในการประชุมกลุ่ม

ข้อเสนอแนะ

ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1. การกำหนดมาตรฐานด้วยวิธีการบูคมาร์ค เหมาะสำหรับการกำหนดมาตรฐานที่ผู้สอบต้องนำผลการสอบไปตัดสินอนาคตของผู้สอบ เนื่องจากต้องการคะแนนจุดตัดที่เป็นมาตรฐานเดียวกันหมด และเป็นคะแนนจุดตัดที่มีคุณภาพสูง วิธีนี้ให้ค่าความแปรปรวนของผู้ตัดสินไม่มาก จึงทำให้มีคะแนนจุดตัดที่คงเส้นคงวา นอกจากนี้ยังเหมาะสำหรับแบบสอบที่มีข้อสอบจำนวนมาก เช่น 100 ข้อขึ้นไป มีลักษณะการให้คะแนนแบบหลายค่า และให้คะแนนแบบสองค่าอยู่ในฉบับเดียวกัน และเหมาะสำหรับคะแนนจุดตัดหลายระดับเพราะสะดวกในการกำหนดคะแนนจุดตัดเพียงคั่นหนังสือในหน้าที่คิดว่าผู้สอบคาบเส้นระดับนั้นสามารถทำได้และหน้าถัดไปไม่สามารถทำได้

สำหรับการสอบที่ตัดสินอนาคตผู้สอบ อาทิเช่น การทดสอบระดับชาติเพื่อให้ผลย้อนกลับถึงผู้เรียนว่า ผู้เรียนอยู่ในระดับใด เช่น อาจแบ่งเป็น 4 ระดับคือ ระดับดีมาก ระดับดี ระดับพอใช้ และระดับตก หรือ การสอบใบประกอบวิชาชีพต่าง ๆ เช่น อาจแบ่งเป็น 4 ระดับคือ ระดับเชี่ยวชาญ ระดับชำนาญการ ระดับพื้นฐาน และระดับตก หรือการทดสอบในห้องเรียนที่ต้องนำผลการเรียนมาใช้พิจารณาการเข้าเรียนต่อระดับอุดมศึกษา หรือเพื่อสอบชิงทุน

ข้อเสนอแนะในการวิจัย

1. เนื่องจากงานวิจัยนี้ยังไม่มีมีการตรวจสอบคุณภาพมาตรฐานหรือคะแนนจุดตัดว่ามีความตรงหรือไม่ จึงควรมีการเปรียบเทียบคุณภาพของการกำหนดคะแนนจุดตัดด้วยวิธีบูคมาร์ค (Bookmark Method) กับวิธีอื่นเช่น Single-Passage Bookmark Method, Contrasting Groups, Jaeger – Mills ซึ่งเป็นการหาความตรงภายนอก นั่นคือเปรียบเทียบคะแนนจุดตัดวิธีบูคมาร์คกับวิธีอื่นว่ามีความสอดคล้องของคะแนนจุดตัดมากน้อยเพียงใด ส่วนความตรงภายในสามารถศึกษาจาก generalizability theory เพื่อประมาณค่า variance components

2. วิธีการกำหนดมาตรฐานหรือคะแนนจุดตัดแนวใหม่ ที่มีลักษณะแบบสอบเป็นการให้คะแนนแบบสองค่า(เช่น แบบเลือกตอบ) และมากกว่าสองค่า(เช่นอัตนัย) ในฉบับเดียวกัน หรือเป็นแบบสอบแบบใดแบบหนึ่ง เช่น วิธีแองกอฟแบบปรับขยาย (Extended Angoff) (Hambleton & Plake, 1995; Impara & Plake, 1997; Brandon, 2004) วิธีการเส้นภาพเด่น (Dominant Profile) (Putnam, Pence & Jaeger, 1995; Plake, Hambleton & Jaeger, 1997) วิธีการใช้การตัดสินเชิงนโยบาย (Judgmental Policy Capturing) (Jaeger, 1995) ซึ่งยังไม่ได้ศึกษาวิธีการดังกล่าว จึงควรทำวิจัยเกี่ยวกับการกำหนดคะแนนจุดตัดด้วยวิธีดังกล่าว

3. งานวิจัยนี้ยังไม่ได้ศึกษาถึงความคิดเห็นของผู้ตัดสินใจในการใช้การกำหนดมาตรฐานด้วยวิธีบูคมาร์ค จึงควรทำวิจัยเชิงคุณภาพกับผู้ตัดสินใจหลังจากกำหนดมาตรฐานด้วยวิธีบูคมาร์คกับวิธีอื่น ๆ เพื่อตรวจสอบประสิทธิผล และกระบวนการคิดในการกำหนดคะแนนจุดตัด ความเหมาะสมและความเป็นไปได้ของวิธีการดังกล่าว เพื่อนำผลที่ได้ไปใช้ปรับปรุงกระบวนการกำหนดคะแนนจุดตัดต่อไป