

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

องค์ประกอบของกระบวนการศึกษาประกอบด้วยวัตถุประสงค์ทางการศึกษา (Objective) การจัดประสบการณ์การเรียนรู้ (Learning Experience) และวิธีการวัดและประเมินผล (Measurement, Evaluation) ซึ่งทั้ง 3 องค์ประกอบต่างมีความสัมพันธ์กัน นั่นคือ การจัดประสบการณ์เรียนรู้ และการวัดและประเมินผลต้องใช้วัตถุประสงค์ทางการศึกษานำทาง เพื่อให้นักเรียนบรรลุตามวัตถุประสงค์ที่กำหนดไว้ ขณะเดียวกันสารสนเทศจากการวัดและประเมินจะเป็นข้อมูลย้อนกลับเพื่อปรับปรุงการเรียนการสอน และการกำหนดวัตถุประสงค์ให้เหมาะสมต่อไป จะเห็นว่าการวัดและประเมินผลมีบทบาทสำคัญต่อกระบวนการจัดการเรียนรู้ของผู้เรียน เพราะการวัดและประเมินผลจะให้สารสนเทศเพื่อพัฒนาผู้เรียนว่าบรรลุตามมาตรฐานที่กำหนดหรือไม่เพียงไร

สำหรับการจัดการเรียนการสอนของคณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา ได้บรรจุรายวิชาการวัดและประเมินผลทางการศึกษาเป็นรายวิชาบังคับ ในหลักสูตรปริญญาตรีทุกสาขาวิชาเพื่อให้สอดคล้องกับมาตรฐานวิชาชีพครู โดยใช้ชื่อวิชา 400204 การวัดและประเมินในชั้นเรียน ซึ่งในแต่ละปี คณะศึกษาศาสตร์ต้องเปิดกลุ่มรายวิชาการวัดและประเมินผลการศึกษาประมาณ 15 - 20 กลุ่ม เพื่อให้การเรียนการสอนและการวัดและประเมินไปในแนวทางเดียวกัน จึงจำเป็นต้องหาวิธีการวัดและประเมินผลที่เป็นมาตรฐาน ผู้วิจัยซึ่งเป็นหนึ่งในฐานะผู้สอนในรายวิชานี้ เล็งเห็นว่าวิธีการหนึ่งคือ การมีแบบสอบวัดผลสัมฤทธิ์ทางการเรียนที่เป็นมาตรฐานในรายวิชาการวัดและประเมินผลการศึกษา ผู้วิจัยจึงมีความสนใจที่จะพัฒนาแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาการวัดและประเมินผลในชั้นเรียน เพื่อให้ได้แบบสอบมาตรฐานที่สามารถวัดนิสิตได้ใกล้เคียงกับความสามารถที่แท้จริง นอกจากนี้แล้วควรมีมาตรฐาน หรือคะแนนจุดตัดของแบบสอบที่มีคุณภาพเช่นเดียวกัน

ความเป็นมาของการกำหนดมาตรฐาน เริ่มมาจากการทดสอบแบบอิงเกณฑ์ที่มีมาตั้งแต่ ค.ศ.1951 เมื่อฟลานาแกน (Flanagan) ได้อธิบายคำว่า “เกณฑ์ปกติ” และ “มาตรฐาน” ฟลานาแกนอธิบายถึงการทดสอบแบบอิงเกณฑ์ในลักษณะที่ผู้เรียนจะแปลหนังสือได้โดยผิดไม่เกิน 3 แห่ง อีเบล (Ebel,1962) ก็มีแนวคิดเช่นเดียวกันว่า การทดสอบใด ๆ ก็ตามจะมีความหมายอย่างสมบูรณ์ก็ต่อเมื่อผลการสอบนั้นต้องบอกได้ว่า ผู้สอบมีความรู้อยู่ในระดับใดของเนื้อหาทั้งหมดที่

ทำการสอบ ในปี ค.ศ. 1962 เกลเซอร์และเคลาส์ (Glaser and Klaus) เป็นผู้ใช้คำว่า การวัดผลแบบอิงเกณฑ์ (Criterion-Reference Measurement) เป็นครั้งแรก (Berk, 1980 อ้างถึงใน สุรินทร์ แสงจันทร์, 2528)

การศึกษาค้นคว้าในช่วงแรก ๆ การใช้คะแนนจุดตัด มักใช้การกำหนดขึ้นจากการสร้างระดับเกณฑ์ที่สูง ๆ ไว้ และใช้เทคนิคแบบธรรมดาในการกำหนดระดับเกณฑ์ขึ้นมา เป็นต้นว่า การกำหนดมาตรฐานความรอบรู้โดยใช้เปอร์เซ็นต์การตอบถูก 80% 85% 90% หรือ 100% การกำหนดระดับเกณฑ์ดังกล่าวมักจะใช้การคาดคะเนจากความรู้สึก โดยไม่คำนึงถึงสถานการณ์เงื่อนไขอื่น ๆ ที่อาจแทรกซ้อนเข้ามาตลอดจนคุณสมบัติของข้อสอบ เช่น ระดับความยากที่มีผลต่อการกำหนดระดับเกณฑ์เป็นอย่างมาก

การกำหนดระดับเกณฑ์ในยุคเริ่มต้น เกิดจากการเคลื่อนไหวของการเขียนจุดประสงค์การสอนในรูปจุดประสงค์เชิงพฤติกรรม เมเกอร์ (Mager, 1962 อ้างถึงใน บุญเชิด ภิญญอนันตพงษ์, 2526) เป็นผู้นำในการเขียนจุดประสงค์เชิงพฤติกรรม ได้เสนอการกำหนดมาตรฐานโดยเขียนเป็นองค์ประกอบหนึ่งของจุดประสงค์เชิงพฤติกรรม ตัวอย่างเช่น

“นักเรียนต้องสามารถแก้สมการเชิงเส้นได้ถูกต้องภายในเวลา 30 นาที”

จากตัวอย่างจะเห็นว่า การกำหนดระดับเกณฑ์ว่า “ถูกต้องภายใน 30 นาที” เป็นการกำหนดมาตรฐานการปฏิบัติที่ใช้ภาษาแต่ไม่ใช่ค่าปริมาณที่แท้จริง

นักวัดผลอีกท่านหนึ่งคือ บลูม (Bloom, 1968 อ้างถึงใน บุญเชิด ภิญญอนันตพงษ์, 2526) มีส่วนเข้ามาเกี่ยวข้องกับการพัฒนาการกำหนดมาตรฐาน โดยเสนอแนวคิดเกี่ยวกับการเรียนการสอนเพื่อทราบว่า “ถ้าให้เวลาอย่างเพียงพอ และให้ความช่วยเหลือที่เหมาะสมแล้ว นักเรียน 95% ของทั้งหมดจะสามารถเรียนรู้เนื้อหาวิชาได้ถึงระดับความรอบรู้ที่กำหนดไว้ได้” อย่างไรก็ตามแนวความคิดของบลูมก็ไม่ได้ให้วิธีการกำหนดระดับความรอบรู้ที่ชัดเจนไว้

ผู้นำทางการทดสอบแบบอิงเกณฑ์อีกท่านคือ โพแฮม (Popham, 1969) ได้ยืนยันซ้ำถึงความสำคัญของมาตรฐานการปฏิบัติว่า “การสร้างมาตรฐานการปฏิบัติโดยกำหนดเป็นระดับผลสัมฤทธิ์ของนักเรียนชั้นต่ำก่อนที่จะทำการสอน จะเป็นประโยชน์อย่างมากต่อการวางแผนประเมินผลการสอน” และได้ยกตัวอย่างคล้ายเมเกอร์ดังนี้

“นักเรียนต้องสามารถแก้โจทย์ปัญหาเกี่ยวกับเส้นรอบรูปได้ถูกต้อง 10 ใน 15 ข้อ”

นอกจากนี้ยังมีนักประเมินผลอีกท่านหนึ่งคือ ไทเลอร์ (Tyler, 1973 อ้างถึงใน บุญเชิด ภิญญอนันตพงษ์, 2526) ได้เน้นถึงการกำหนดมาตรฐานในการชี้บอกระดับความรู้ตามโปรแกรม และได้ให้ตัวอย่างไว้ดังนี้

“นักเรียนที่แสดงว่ามีความรอบรู้ในเนื้อหาวิชาเฉพาะที่เรียนรู้ไปแล้วจะต้องมีความสามารถในการตอบปัญหาที่กำหนดให้ได้ถูกต้อง 85% ของทั้งหมด ภายในเวลาที่กำหนด”

การนำเอาระดับเกณฑ์ ระบุไว้ในจุดประสงค์การสอนดังกล่าว เป็นการกำหนดเกณฑ์โดยใช้ภาษาธรรมดาซึ่งยังไม่เป็นค่าเชิงปริมาณที่แท้จริง และลักษณะการกำหนดระดับเกณฑ์มักจะใช้ประสบการณ์ในการคาดคะเน และเป็นการกำหนดตามประเพณีนิยมมากกว่าจะเป็นกระบวนการทางวิทยาศาสตร์ จึงได้มีวิธีการกำหนดค่าเกณฑ์ในรูปคะแนนจุดตัดที่เป็นระบบขึ้นมา

วิธีการในการกำหนดคะแนนจุดตัดมีด้วยกันหลายวิธี เช่น วิธีการนับลดจาก 100% (Counting Backwards from 100%) ซึ่งกลาส (Glass, 1978) เสนอไว้โดยให้ผู้เชี่ยวชาญหรือครูผู้สอนเป็นผู้พิจารณาข้อสอบและจุดประสงค์เชิงพฤติกรรมว่า จุดประสงค์เชิงพฤติกรรมแต่ละข้อมีความสำคัญมากน้อยเพียงใด แล้วพิจารณากำหนดคะแนนจุดตัดของแบบทดสอบจากเกณฑ์ที่ตั้งไว้ ซึ่งแล้วแต่ว่าผู้ตั้งเกณฑ์จะลดลงมาที่เปอร์เซ็นต์ วิธีการนี้เป็นการกำหนดเกณฑ์ที่ค่อนข้างเป็นอัตนัย เนื่องจากใช้ความคิดเห็นส่วนบุคคลในการพิจารณา ต่อมาจึงมีผู้พัฒนาวิธีการกำหนดคะแนนจุดตัดโดยการพิจารณาเนื้อหาวิชาของแบบสอบเป็นรายข้อ แล้วระบุว่าผู้สอบที่มีความสามารถหรือสมรรถภาพขั้นต่ำที่จะยอมรับว่าผ่านได้นั้นควรทำแบบสอบได้เป็นจำนวนเท่าไร ก็จะทำให้ได้คะแนนจุดตัดที่เหมาะสมขึ้น ซึ่งวิธีการดังกล่าวนี้คือ การพิจารณาตัดสินจากความสามารถต่ำสุดของผู้สอบ (Judging Minimal Competence) ตามที่กลาส (Glass, 1978) เสนอไว้ หรือเป็นวิธีการพิจารณาจากข้อคำถามในแบบสอบตามที่ลิฟวิงตัน และซีกี (Livingston & Zieky, 1982) เสนอไว้ วิธีการนี้มีหลายเทคนิคด้วยกันคือ เทคนิคของแองกอฟ (Angoff) เทคนิคของอีเบล (Ebel) และเทคนิคของนีเดลสกี (Nedelsky) ทั้ง 3 เทคนิคต่างก็มีวิธีการที่เหมือนกันคือ ผู้เชี่ยวชาญจะต้องพิจารณาถึงความรู้และทักษะของผู้สอบที่อยู่ในระดับต่ำสุดหรือที่เรียกว่า ระดับคาบเส้น (Borderline) ว่าสมควรให้ผ่านหรือไม่ ซึ่งเป็นเทคนิคที่สะดวกแก่ผู้ใช้มากเนื่องจากสามารถกำหนดคะแนนจุดตัดได้ทั้งก่อนการสอบ และหลังการสอบเสร็จสิ้นแล้ว โดยที่ผู้เชี่ยวชาญจะต้องพิจารณาตัดสินข้อความในแบบทดสอบร่วมกัน

จะเห็นว่าเทคนิคทั้ง 3 ในทางทฤษฎีเป็นสิ่งที่ดี แต่ในทางปฏิบัติยังมีปัญหา คือ 1) ผู้เชี่ยวชาญแต่ละคนมีความคิดที่แตกต่างกันเกี่ยวกับความสามารถต่ำสุดของผู้สอบ เพราะฉะนั้นคะแนนจุดตัดที่ผู้เชี่ยวชาญกำหนดจะสะท้อนถึงความแตกต่างของผู้สอบด้วย

2) การประมาณความสามารถต่ำสุดของผู้สอบจากการพิจารณาข้อสอบจำนวนมากเช่น จำนวน 50 ข้อ นับว่าเป็นงานที่ยาก เพราะว่าผู้สอบบางคนจะได้คะแนนสูงในบางข้อ และบางคนได้คะแนนต่ำในบางข้อ เพราะฉะนั้นการคาดคะเนจากความสามารถต่ำสุดจึงเป็นสิ่งที่เป็นไปได้ แต่ถ้าผู้สอบที่อยู่ในระดับความสามารถต่ำสุดอยู่ในระดับเดียวกัน การประมาณความสามารถต่ำสุดของผู้สอบ จากการทำแบบสอบทั้งฉบับจะทำได้ง่ายกว่าการพิจารณาแต่ละข้อคำถาม (Rowley, 1982) ดังนั้น ลิเวลตัน และซีกี (Livingston & Zieky, 1982) จึงเสนอวิธีการกำหนดจุดตัดวิธีใหม่คือ การใช้วิธีกลุ่มตรงกันข้าม (Contrasting Group Method) ซึ่งเป็นวิธีการตัดสินที่ขึ้นอยู่กับผู้สอบมากกว่าการตัดสินเกี่ยวกับข้อคำถามในแบบทดสอบ วิธีการนี้มีข้อดีคือ ผู้ตัดสินจะคุ้นเคยกับการพิจารณาทักษะของผู้เรียน และสามารถพิจารณาได้อย่างมีความหมาย นอกจากนี้ข้อสังเกตที่ได้จากวิธีการนี้จะสอดคล้องกับสภาพความเป็นจริงมากกว่าการพิจารณาจากข้อคำถามในแบบทดสอบ การใช้วิธีกลุ่มตรงกันข้ามในการกำหนดคะแนนจุดตัดมีหลายวิธี วิธีที่ซับซ้อนน้อยที่สุดคือ การเขียนกราฟแสดงการแจกแจงของคะแนนของกลุ่มรอบรู้ และกลุ่มไม่รอบรู้ แล้วพิจารณาเส้นกราฟของผู้สอบทั้งสองกลุ่มตัดกัน ณ จุดใด จุดนั้นก็คือ คะแนนจุดตัด (Mills, 1983)

ในการใช้วิธีกลุ่มตรงกันข้ามเพื่อกำหนดคะแนนจุดตัดนั้นปัญหาที่พบก็คือ ไม่สามารถแบ่งได้อย่างแน่ชัดระหว่างผู้ที่รอบรู้กับผู้ที่ไม่รอบรู้ ดังนั้น ลิเวลตัน และซีกี (Livingston & Zieky, 1982) ได้เสนอวิธีการกำหนดจุดตัดอีกวิธีหนึ่งคือ วิธีกลุ่มคาบเส้น (Borderline Group Method) ซึ่งวิธีนี้มาจากแนวคิดที่ว่า มาตรฐานจะเป็นคะแนนที่ได้จากผู้สอบซึ่งมีความสามารถอยู่ในระดับคาบเส้น วิธีการนี้แทนที่จะให้ผู้เชี่ยวชาญเดาถึงความสามารถของผู้สอบที่อยู่ในระดับคาบเส้นก็ใช้วิธีการให้ผู้เชี่ยวชาญระบุความรู้ความสามารถของผู้สอบที่อยู่ในระดับคาบเส้น และพิจารณาคะแนนจุดตัดจากผู้สอบดังกล่าว มาตรฐานที่กำหนดสำหรับการผ่านจะใช้มัธยฐานของคะแนนผู้สอบที่อยู่ในกลุ่มคาบเส้น ประโยชน์สำคัญของวิธีการนี้คือง่ายในการนำไปใช้และง่ายในการอธิบาย (กัญจนา ลินทรัตน์ศิริกุล, 2534)

เมื่อเข้าสู่ ค.ศ.1990 ในอเมริกาได้เกิดแนวโน้มจากการประเมินระดับประเทศ (Large Scale) 2 วิธีคือ 1) มีการใช้ข้อสอบที่ให้คะแนนมากกว่า 2 ค่า เพิ่มมากขึ้น 2) มีความต้องการคะแนนจุดตัดที่หลากหลายระดับ และแล้วในเดือนกรกฎาคม ปี ค.ศ.1995 แนวโน้มดังกล่าวได้รับการยืนยันโดยข้อบังคับแห่งชาติสำหรับการดำเนินโปรแกรมที่อยู่ภายใต้ Title I of Elementary and Secondary Education Act of 1965 และได้รับการเปลี่ยนแปลงแก้ไขโดย The Improving America's School Act of 1994 (U.S. Department of Education, 1995)

cited in Berk, 1996) ข้อบังคับดังกล่าวต้องการให้มลรัฐต่าง ๆ ได้ใช้วิธีการที่หลากหลายในการประเมินซึ่งประกอบด้วยการประเมินทักษะการคิดที่ซับซ้อนและความเข้าใจในเนื้อหา และอธิบายระดับการปฏิบัติหรือจุดตัด 3 ระดับคือ ระดับขั้นสูง (Advanced) ระดับชำนาญการ (Proficient) และระดับชำนาญการบางส่วน (Partially Proficient) ในการประเมินนักเรียนทุกคน Berk (1996) และ Zieky (2001) ได้กล่าวว่า ได้เกิดสิ่งที่น่าสนใจในการทำมาตฐานที่จะนำมาตฐานไปใช้กับแบบสอบเขียนตอบ แบบสอบภาคปฏิบัติ และการทดสอบปรับเหมาะในคอมพิวเตอร์ นับจากนั้นจึงมีงานวิจัยเกี่ยวกับวิธีการกำหนดมาตรฐานแนวใหม่ออกมาเป็นจำนวนมาก เช่น **วิธีแองกอฟแบบปรับขยาย** (Extended Angoff) (Hambleton & Plake, 1995; Impara & Plake, 1997; Goodwin, 1999; Brandon, 2004) **วิธีการบุ้คมาร์ค** (Bookmark) (Kiplinger, 1997; Huynh, 1998; Zwick, Senturk, Wang & Barbara, Loomis, 2001; Van der Schoot, 2002; Buckendahl, Smith, Impara & Plake, 2002; Wang, 2003; Green, Trimble & Lewis, 2003; Beretvas, 2004) **วิธีการเส้นภาพเด่น** (Dominant Profile) (Putnam, Pence & Jaeger, 1995; Plake, Hambleton & Jaeger, 1997) **วิธีการใช้การตัดสินเชิงนโยบาย** (Judgmental Policy Capturing) (Jaeger, 1995) เป็นต้น

วิธีแองกอฟแบบปรับขยาย วิธีนี้ถูกพัฒนาสำหรับการประเมินการปฏิบัติที่ให้คะแนนมากกว่า 2 ค่า โดยกำหนดให้คณะผู้ตัดสินทำการประมาณคะแนนคาดหวังของผู้สอบที่คาบเส้นหรือผู้สอบที่เกือบไม่ได้รับการรับรอง (Borderline Candidates) ในแบบฝึกหัดแต่ละฉบับที่ให้คะแนนมากกว่า 2 ค่า งานของผู้ตัดสินก็คือ กำหนดความน่าจะเป็น (เช่น คะแนนที่คาดหวัง) ที่ผู้สอบคาบเส้นสามารถตอบข้อสอบถูก คะแนนคาดหวังจากแบบฝึกหัดของผู้ตัดสินแต่ละคนสามารถนำไปใช้ในมาตรฐานการปฏิบัติในชุดของแบบฝึกหัด จากนั้นมาตรฐานของผู้ตัดสินจะถูกนำมาเฉลี่ยเพื่อให้ได้มาตรฐานตามที่ต้องการสำหรับการใช้ในการรับรอง วิธีการแองกอฟแบบปรับขยายเป็นวิธีที่ง่ายมีความชัดเจน กระบวนการกำหนดน้ำหนักทำให้ผู้ตัดสินกำหนดค่าของคำถามแต่ละข้อแตกต่างกัน ซึ่งการกำหนดน้ำหนักมากในบางคำถามแสดงว่า ผู้ตัดสินให้ความสำคัญกับข้อนั้นมาก อย่างไรก็ตาม วิธีการแองกอฟแบบปรับขยายอาจใช้ไม่ได้เมื่อนำไปใช้กับการปฏิบัติที่เป็นธรรมชาติซึ่งมีความยืดหยุ่นสูง

วิธีการบุ้คมาร์ค วิธีการบุ้คมาร์คเป็นการรวมเนื้อหาข้อสอบกับผลการสอบของผู้สอบเข้าด้วยกัน โดยมีคู่มือที่มีการเรียงข้อสอบจากข้อง่ายไปข้อยาก (Mitzel, Lewis, Patz, & Green, 2001 cite in Beretvas, 2004) ซึ่งข้อสอบที่ถูกจัดเรียงตามค่าความยากจากข้อสอบง่ายสุดไปยังข้อสอบที่ยากสุด จะใช้กระบวนการ IRT ในการวิเคราะห์เพื่อให้ทราบถึงค่าความยากของข้อสอบ

และความสามารถของนักเรียนที่อยู่บนมาตรเดียวกัน (Same Continuum) ข้อสอบแบบเขียนตอบ ก็จะถูกนำมาไว้ในมาตรเดียวกับข้อสอบแบบเลือกตอบโดยคะแนนแต่ละคะแนนในข้อสอบแบบเขียนตอบจะถูกนำมากำหนดบนมาตร แล้วจัดเรียงลำดับลงในคู่มือ คู่มือที่ได้จะนำมาให้ผู้ตัดสิน พิจารณาประกอบการหาคะแนนจุดตัด โดยผู้ตัดสินจะพิจารณาข้อสอบทีละข้อ จากข้อง่ายไปยัง ข้อยาก โดยให้ผู้ตัดสินตอบคำถามว่า “ผู้สอบที่คาบเส้นมีโอกาสตอบข้อสอบข้อนั้นถูก 67% หรือไม่” พิจารณาทีละข้อไปเรื่อย ๆ จนถึงข้อที่ผู้สอบกลุ่มคาบเส้นไม่มีโอกาสตอบถูก ก็จะใช้ที่คั่น หนังสือคั่นตรงหน้านั้นไว้ จากนั้นนำผลการตัดสินของแต่ละคนมาอภิปรายร่วมกัน เปิดโอกาสให้ แต่ละคนหาคะแนนจุดตัดอีกครั้งแล้วจึงนำมาหาค่าเฉลี่ย จะเห็นได้ว่าวิธีนี้ ผู้ตัดสินจะพิจารณา แบบสอบในภาพรวมโดยอยู่บนพื้นฐานการพิจารณาจากผลคะแนนของนักเรียน นอกจากนี้วิธี บัญญัติก็ยังถูกออกแบบมาให้ใช้กับข้อสอบแบบเลือกตอบที่ให้คะแนนแบบ 2 ค่า และข้อสอบแบบ เขียนตอบที่ให้คะแนนมากกว่า 2 ค่า อย่างเท่าเทียมกันเพราะข้อสอบทุกข้ออยู่บนสเกล IRT เหมือนกัน อย่างไรก็ตามวิธีนี้ควรมีข้อสอบจำนวนมากพอ สำหรับพิจารณานักเรียนที่อยู่ใน มาตรฐานระดับก้าวหน้า หรือระดับตก ซึ่งทั้งสองระดับเป็นระดับสุดโต่ง (เก่งมาก กับอ่อนมาก) ซึ่งอาจทำให้เกิด floor and ceiling effect ได้

วิธีการเส้นภาพเด่น (Dominant Profile Method) วิธีนี้จะใช้ยุทธวิธีสร้างฉันทมติ (Consensus-Building) ผู้ตัดสินจะถูกถามเพื่อให้เกิดกฎที่เป็นนโยบาย (ตัดสินใจ) เพื่อใช้กำหนด คะแนนงาน หรือ แบบฝึกหัดที่ใช้ประกอบการประเมินเป็นค่าต่ำสุดที่ต้องการให้ผ่าน องค์ประกอบการ ทดแทน (Compensatory) และองค์ประกอบที่เชื่อมโยง (Conjunctive) มีความยืดหยุ่นสำหรับ กฎการตัดสินใจ ดังนั้นผู้ตัดสินสามารถสร้างกฎการตัดสินใจที่มีลักษณะซับซ้อนเพื่อให้มีความ หมายถึงผู้สอบที่มีลักษณะคุณภาพผ่านการประเมิน วิธีนี้มีความซับซ้อนในการใช้มากกว่า มาตรฐานการปฏิบัติที่มีเพียงมาตรฐานเดียว

วิธีการใช้การตัดสินเชิงนโยบาย (Judgmental Policy Capturing Method) วิธีการ คือ คณะผู้ตัดสินต้องได้รับการฝึกอย่างเข้มข้นเพื่อให้เข้าใจความหมายของแต่ละคะแนนในแต่ละ แบบฝึกหัด เมื่อแสดงเส้นภาพของคะแนนนักเรียนจากทุกแบบฝึกหัดแล้ว ผู้ตัดสินก็จะสร้างรูปภาพ ทางความคิดของคุณภาพผู้สอบด้านการปฏิบัติ คะแนนเส้นภาพเหล่านี้จะถูกนำมาจำแนกเพื่อ กำหนดเป็นคะแนนเส้นภาพลงในแต่ละประเภทที่ระบุข้อความมาตรฐานการปฏิบัตินั้น ๆ (เช่น ต่ำ มาก ค่อนข้างต่ำ ต่ำเล็กน้อย สูงเล็กน้อย ค่อนข้างสูง สูงมาก) ผลการตัดสินจะถูกนำมาวิเคราะห์ โดยแต่ละประเภทจะถูกนำมากำหนดเป็นค่าตัวเลข กระบวนการทางสถิติ (เช่น การวิเคราะห์การ ถดถอยพหุ) จะถูกนำมาใช้เพื่อกำหนดความสัมพันธ์ทางคณิตศาสตร์ระหว่างเส้นภาพคะแนน ผู้สอบ กับประเภทการปฏิบัติ (Performance Categories) ความสัมพันธ์ดังกล่าวจะกำหนด

น้ำหนักเพื่อนำไปใช้กับคะแนนของผู้สอบในแต่ละแบบฝึกหัดและสำหรับคำนวณคะแนนการปฏิบัติในผู้สอบแต่ละคน คะแนนการปฏิบัติทั้งหมดจะถูกคำนวณสำหรับผู้สอบแต่ละคน เส้นภาพที่เป็นคะแนนแบบฝึกหัดจะนำมาแสดงให้ผู้ตัดสินพิจารณา เส้นภาพเหล่านี้ก็จะถูกจัดเรียงลำดับคะแนนการปฏิบัติสูงสุดไปยังคะแนนการปฏิบัติต่ำสุด ผู้ตัดสินจะถูกถามเพื่อกำหนดคะแนนการปฏิบัติต่ำสุดที่สามารถประเมินผ่านได้ ข้อเสนอแนะจากผู้ตัดสินทั้งหมดจะถูกนำมาจัดเป็นตารางการแจกแจงค่ามัธยฐานจะถูกนำมากำหนดเป็นมาตรฐานการปฏิบัติ ขั้นตอนทั้งสองจะถูกนำมาทำซ้ำโดยให้โอกาสผู้ตัดสินอธิบายเหตุผลจากข้อเสนอแนะเดิมของพวกเขา มาตรฐานการปฏิบัติขั้นสุดท้ายจะยึดเสียงข้างมากของผู้ตัดสิน วิธีนี้ใช้เวลาในการฝึกผู้ตัดสินเป็นระยะเวลานานและเมื่อใช้เวลาฝึกนาน ค่าใช้จ่ายก็จะสูงตามไปด้วย

จากที่กล่าวมาจะเห็นว่า วิธีการกำหนดมาตรฐานส่วนใหญ่ยังคงเป็นการตัดสินที่ใช้ดุลยพินิจของผู้ตัดสินซึ่งเป็นข้อจำกัดที่สำคัญ จึงได้มีความพยายามพัฒนาปรับปรุงวิธีการกำหนดมาตรฐานขึ้นมาเพื่อลดข้อจำกัดเกี่ยวกับตัวผู้ตัดสินเช่น การให้ข้อมูลสารสนเทศประกอบการตัดสินที่เพียงพอ การจัดอภิปรายกลุ่ม การดำเนินการตัดสินหลายรอบ เป็นต้น

ปัจจุบันวิธีการกำหนดมาตรฐานที่อาศัยดุลยพินิจของผู้ตัดสินที่ได้รับความนิยมใช้กันอย่างกว้างขวางคือ วิธีการกำหนดคะแนนด้วยวิธีแองกอฟ แต่วิธีนี้เหมาะสำหรับแบบสอบที่เป็นแบบหลายตัวเลือก ดังนั้นถ้าเป็นแบบสอบแบบเลือกหลายคำตอบ และมีแบบอัตนัยหรือเขียนตอบร่วมอยู่ด้วยในฉบับเดียวกัน วิธีการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟจะไม่เหมาะสม เพราะการให้คะแนนไม่เป็นระบบถูกให้ 1 ผิดให้ 0 แบบสอบแบบดังกล่าวควรใช้วิธีการกำหนดคะแนนจุดตัดแนวใหม่ด้วยวิธีบุ๊กมาร์ค (Bookmark) เนื่องจากในสหรัฐอเมริกาวิธีนี้ได้รับความนิยมใช้กันอย่างกว้างขวางใน 28 มลรัฐ (Egan, 2001 cited in Beretvas, 2004) แต่ในประเทศไทยวิธีดังกล่าวยังไม่ค่อยถูกนำมาใช้กันมากนัก วิธีนี้มีจุดเด่นคือ ช่วยจัดระบบการคิดแก่ผู้ตัดสินทำให้ผู้ตัดสินตัดสินได้ง่ายขึ้น เนื่องจากมีการจัดเรียงข้อสอบที่เป็นระบบในรูปของคู่มือ และเหมาะสำหรับแบบสอบที่มีการให้คะแนนแบบ 0, 1 (เช่น แบบเลือกตอบ) และไม่ใช่ 0, 1 (เช่น แบบอัตนัย) โดยงานวิจัยของสุริพร อนุศาสนนันท์ (2550) ได้เปรียบเทียบคุณภาพมาตรฐานระหว่างวิธีแองกอฟที่ได้รับการปรับปรุงกับวิธีบุ๊กมาร์ค แต่งานวิจัยงานนั้นวิธีกำหนดมาตรฐานด้วยวิธีบุ๊กมาร์คใช้กับแบบสอบที่มีระบบการให้คะแนนเป็นแบบ 0, 1 ยังไม่ได้ศึกษาการกำหนดมาตรฐานบุ๊กมาร์คที่กำหนดในแบบสอบที่มีการให้คะแนนแบบ 0, 1 และไม่ใช่ 0, 1 ในฉบับเดียวกัน ซึ่งการคำนวณค่าความสามารถจะใช้สูตรต่างกัน รวมทั้งการวิเคราะห์ข้อสอบโดยทฤษฎีการตอบสนองข้อสอบนอกจากใช้โปรแกรม Xcalibre สำหรับข้อสอบเลือกตอบ แล้วยังใช้โปรแกรม Multilog สำหรับข้อสอบอัตนัย ด้วยเหตุนี้ผู้วิจัยจึงมีความสนใจที่พัฒนาแบบสอบวัดผลสัมฤทธิ์ทางการ

เรียนวิชาการวัดและประเมินผลการศึกษา รวมทั้งกำหนดมาตรฐาน โดยวิธีบ្រុคมาร์คในแบบสอบที่พัฒนาขึ้นที่มีลักษณะเป็นแบบเลือกตอบ และอัตนัย โดยการวิจัยนี้ให้ผู้ตัดสินพิจารณาคะแนนจุดตัดของนักเรียนเป็น 7 ระดับ คือ ระดับดีเยี่ยม (A) ระดับดีมาก (B+) ระดับดี (B) ระดับดีพอใช้ (C+) ระดับพอใช้ (C) ระดับอ่อน (D+) และ ระดับอ่อนมาก (D) ซึ่งเป็นลักษณะการกำหนดคะแนนจุดตัดที่มีหลายระดับ งานวิจัยนี้น่าจะเป็นประโยชน์ต่อวงการการศึกษาของไทยโดยเฉพาะการกำหนดมาตรฐานด้วยวิธีบ្រុคมาร์ค ในแบบสอบที่มีลักษณะแบบเลือกตอบ และแบบเขียนตอบที่อยู่ในฉบับเดียวกัน ซึ่งปัจจุบันแบบสอบลักษณะแบบนี้ ได้รับความนิยมน้อย่างกว้างขวาง โดยเฉพาะนำแบบสอบนี้ไปใช้ในการสอบที่ส่งผลต่ออนาคตของผู้สอบ เช่น การทดสอบระดับชาติ การสอบใบประกอบวิชาชีพต่าง ๆ เนื่องจากการสอบเหล่านี้ต้องการคะแนนจุดตัดที่เป็นมาตรฐานเดียวกันหมด และเป็นคะแนนจุดตัดที่มีคุณภาพสูง

วัตถุประสงค์ในการวิจัย

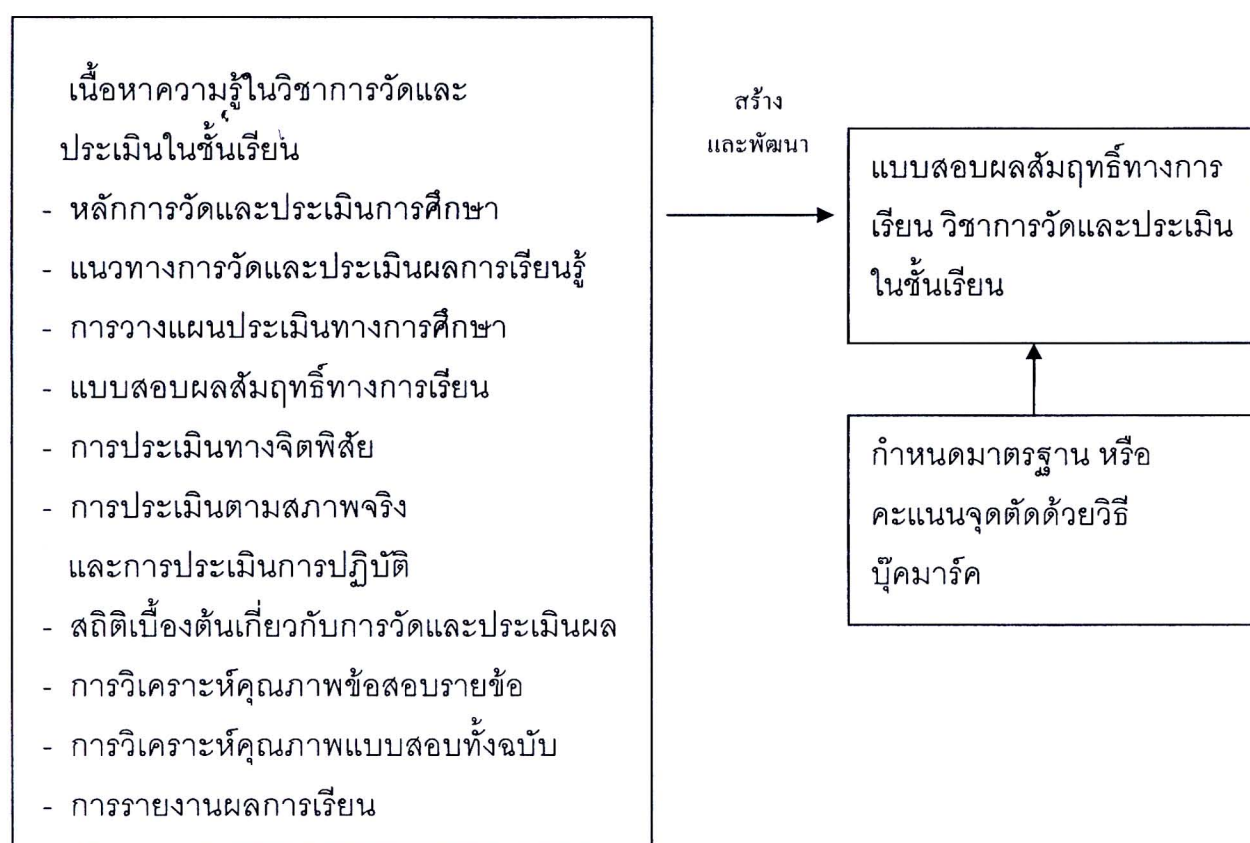
1. เพื่อสร้างแบบสอบผลสัมฤทธิ์ทางการเรียนวิชา การวัดและประเมินผลการศึกษา ระดับปริญญาตรี มหาวิทยาลัยบูรพา
2. เพื่อตรวจสอบความยาก อำนาจจำแนก ความตรง และความเที่ยงของแบบสอบผลสัมฤทธิ์ทางการเรียนวิชา การวัดและประเมินผลในชั้นเรียนที่สร้างขึ้น
3. เพื่อหาคะแนนจุดตัดของแบบสอบที่สร้างขึ้นโดยการกำหนดมาตรฐานด้วยวิธีบ្រុคมาร์ค

ขอบเขตการวิจัย

1. ประชากรที่ใช้ในการวิจัยครั้งนี้ เป็นนิสิตระดับปริญญาตรีที่เรียนรายวิชา 400204 การวัดและประเมินผลการศึกษา ปีการศึกษา 2552 จำนวน 990 คน และเป็นอาจารย์สอนมหาวิทยาลัยบูรพา ระดับปริญญาตรีที่เป็นอาจารย์ประจำ และอาจารย์พิเศษ ที่สอนวิชาวัดและประเมินในชั้นเรียน จำนวน 9 คน
2. เนื้อหาวิชาการวัดและประเมินผลการศึกษา มีดังนี้
 - 2.1 หลักการวัดและประเมินผลการศึกษา
 - 2.2 แนวทางการวัดและประเมินผลการเรียนรู้
 - 2.3 การวางแผนประเมินผลทางการศึกษา
 - 2.4 แบบสอบผลสัมฤทธิ์ทางการเรียน
 - 2.5 การประเมินทางจิตพิสัย

- 2.6 การประเมินตามสภาพจริง และการประเมินการปฏิบัติ
- 2.7 สถิติเบื้องต้นเกี่ยวกับการวัดและประเมินผล
- 2.8 การวิเคราะห์คุณภาพข้อสอบรายข้อ
- 2.9 การวิเคราะห์คุณภาพแบบสอบทั้งฉบับ
- 2.10 การรายงานผลการเรียน

กรอบแนวคิดการวิจัย



ภาพที่ 1-1 กรอบแนวคิดในการวิจัย

คำจำกัดความที่ใช้ในการวิจัย

1. **มาตรฐาน (Standard)** หมายถึง ระดับที่บ่งบอกคุณภาพของนักเรียนในวิชาการวัดและการประเมินผลในการศึกษาคั้งนี้แบ่งนิตีออกเป็น 8 ระดับ คือ ระดับดีเยี่ยม (A) ระดับดีมาก (B+) ระดับดี (B) ระดับดีพอใช้ (C+) ระดับพอใช้ (C) ระดับอ่อน (D+) ระดับอ่อนมาก (D) และระดับตก (F) และในแต่ละระดับจะมีคำอธิบายระดับการปฏิบัติ (Performance Level

Descriptor) ซึ่งเป็นข้อความเพื่อแสดงว่า นิสิตควรมีความรู้ ทักษะ หรือคุณลักษณะในแต่ละระดับอย่างไร

2. **คะแนนจุดตัด (Cut Score)** หมายถึง จุดคะแนนที่ใช้เป็นจุดแบ่งนิสิตในวิชาการวัดและประเมินในชั้นเรียน ตามความสามารถของนิสิต งานวิจัยนี้แบ่งคะแนนจุดตัดออกเป็น 7 จุด คือ 1) จุดคะแนนที่แบ่งความสามารถนิสิตที่มีความสามารถอ่อนมาก ออกจากนิสิตระดับตก 2) จุดคะแนนที่แบ่งนิสิตที่มีความสามารถระดับอ่อนมาก ออกจากนิสิตระดับอ่อน 3) จุดคะแนนที่แบ่งนิสิตที่มีความสามารถระดับอ่อน ออกจากนิสิตระดับพอใช้ 4) จุดคะแนนที่แบ่งนิสิตที่มีความสามารถระดับพอใช้ ออกจากนิสิตระดับดีพอใช้ 5) จุดคะแนนที่แบ่งนิสิตที่มีความสามารถระดับดีพอใช้ ออกจากนิสิตระดับดี 6) จุดคะแนนที่แบ่งนิสิตที่มีความสามารถระดับดี ออกจากนิสิตระดับดีมาก 7) จุดคะแนนที่แบ่งนิสิตที่มีความสามารถระดับดีมาก ออกจากนิสิตระดับดีเยี่ยม

3. **วิธีการกำหนดมาตรฐาน (Standard Setting)** หมายถึง กระบวนการหามาตรฐานเพื่อแบ่งความสามารถของนิสิตออกเป็น 8 ระดับ คือ ระดับดีเยี่ยม (A) ระดับดีมาก (B+) ระดับดี (B) ระดับดีพอใช้ (C+) ระดับพอใช้ (C) ระดับอ่อน (D+) ระดับอ่อนมาก (D) และระดับตก (F) ในงานวิจัยครั้งนี้ใช้วิธีการกำหนดมาตรฐานด้วยวิธีบูคมาร์ค

4. **วิธีการกำหนดมาตรฐานบูคมาร์ค (Bookmark Method)** หมายถึง วิธีการหามาตรฐาน โดยจัดให้มีคู่มือการเรียงข้อสอบจากข้อง่ายสุดไปข้อยากสุดข้อละหนึ่งหน้า มีการวิเคราะห์ข้อสอบเพื่อหาค่าความยาก และความสามารถของผู้สอบในแต่ละข้อโดยนำทฤษฎีการตอบสนองข้อสอบ (IRT) มาใช้ในการวิเคราะห์ คู่มือที่ได้จะนำมาให้ผู้ตัดสินพิจารณาประกอบการหาคะแนนจุดตัด ซึ่งผู้ตัดสินจะพิจารณาทีละข้อ จากข้อง่ายไปยังข้อยาก โดยให้ผู้ตัดสินตอบคำถามว่า ผู้ที่คาบเส้นระดับพอใช้จะมีโอกาสตอบข้อสอบข้อนั้นถูก 67%” พิจารณาทีละข้อไปเรื่อย ๆ จนถึงข้อที่ผู้สอบกลุ่มคาบเส้นระดับพอใช้ไม่มีโอกาสตอบถูก ก็จะใช้ที่คั่นหนังสือคั่นตรงหน้านั้นไว้” ซึ่งเป็นคะแนนจุดตัดระดับพอใช้ (จุดแรก) จากนั้นผู้ตัดสินดำเนินการหาคะแนนจุดตัดระดับดี ระดับดีมาก ระดับดีเยี่ยม ระดับอ่อน ระดับอ่อนมาก และระดับตก ตามลำดับ โดยใช้กระบวนการคล้ายกับที่กล่าวมา จากนั้นนำผลการตัดสินของแต่ละคนมาอภิปรายร่วมกัน แล้วจึงเปิดโอกาสให้ผู้ตัดสินแต่ละคนหาคะแนนจุดตัดอีกครั้ง นำคะแนนจุดตัดเหล่านั้นมาหาค่าเฉลี่ย

5. **ผู้ตัดสิน** หมายถึง อาจารย์สถาบันอุดมศึกษาที่ทำการสอนวิชาการวัดและประเมินผลการศึกษา ระดับปริญญาตรี จำนวน 6 คน โดยการเลือกอย่างเจาะจง (Purposive Sampling) มีคุณสมบัติคือ สอนวิชาการวัดและประเมินผลการศึกษา และจบการศึกษาในระดับ

ปริญญาตรี ปริญญาโท หรือปริญญาเอกในสาขาวิทยาการศึกษา หรือสาขาการวัดและประเมินผล การศึกษา.

ประโยชน์ที่คาดว่าจะได้รับ

1. ได้แบบสอบผลสัมฤทธิ์ทางการเรียนวิชา การวัดและประเมินในชั้นเรียน สำหรับนิสิต ระดับปริญญาตรี คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพาที่มีคุณภาพ
2. ได้คะแนนจุดตัดที่มีคุณภาพ 7 ระดับ เพื่อใช้ในการประเมินผลการเรียนในวิชา 400204 การวัดและประเมินผลในชั้นเรียน
3. ได้ข้อความรู้เกี่ยวกับวิธีการกำหนดมาตรฐานแนวใหม่ คือวิธีการบูคมาร์ค วิธีนี้มี จุดเด่น คือ ช่วยจัดระบบการคิดแก่ผู้ตัดสินใจให้ผู้ตัดสินใจกำหนดคะแนนจุดตัดได้ง่ายขึ้น (McKinley, Newman, & Wiser, 1996) เนื่องจากมีการจัดเรียงข้อสอบเป็นระบบในรูปของคู่มือ นอกจากนี้วิธีการนี้ เหมาะสำหรับการกำหนดมาตรฐานแบบสอบที่ในฉบับเดียวกัน แบบเลือกตอบที่มีการให้คะแนน แบบสองค่าคือ ถูกให้ 1 ผิด ให้ 0 และแบบสอบอัตนัยที่มีการให้คะแนนแบบหลายค่า เช่น 1, 2, 3, 4 คะแนน เป็นต้น
4. เป็นแนวทางในการศึกษาเกี่ยวกับการสร้างและพัฒนาแบบสอบผลสัมฤทธิ์ทางการ เรียน และการกำหนดมาตรฐานในวิชาอื่น ๆ