

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

การศึกษาเป็นกระบวนการที่มุ่งเปลี่ยนแปลงพฤติกรรมของผู้เรียนให้เป็นไปในแนวทางที่พึงประสงค์ กระบวนการศึกษามีองค์ประกอบที่สำคัญ 3 ประการ คือ จุดมุ่งหมายทางการศึกษา (Objectives) การจัดประสบการณ์การเรียนรู้ (Educational Experiences) และวิธีการประเมินผล (Evaluation Procedures) และองค์ประกอบทั้ง 3 ประการนี้ต่างมีความสัมพันธ์กัน โดยที่จุดมุ่งหมายทางการศึกษาจะเป็นตัวชี้แนวทางในการจัดประสบการณ์ในการเรียนการสอน ตลอดจนการเลือกวิธีการประเมินผล การเรียนของผู้เรียนว่าบรรลุเป้าหมายมากน้อยเพียงใด และการประเมินผลจะช่วยให้ข้อมูลย้อนกลับ (Feedback) เกี่ยวกับประสิทธิภาพของการจัดประสบการณ์การเรียนรู้ สอนว่ามีมากน้อยเพียงใด ช่วยให้ครูมีสารสนเทศเพียงพอในการตัดสินใจแก้ไขและปรับปรุงเทคนิคในการจัดกิจกรรมการเรียนรู้ให้เหมาะสมกับความสามารถของนักเรียน และทำให้ครูทราบว่านักเรียนแต่ละคนสามารถสัมฤทธิ์ผลตามวัตถุประสงค์ที่ตั้งไว้หรือไม่อย่างไร (Mehrens and Lamann, 1984)

การศึกษาในระดับประถมศึกษาเป็นการศึกษาขั้นพื้นฐานที่มีจุดมุ่งหมายเพื่อให้ผู้เรียนสามารถพัฒนาคุณภาพชีวิตของตนเองได้ การจัดการศึกษามุ่งปลูกฝังให้ผู้เรียนมีทักษะพื้นฐานในการเรียนรู้ คงสภาพการอ่านออกเขียนได้และคิดคำนวณได้ โครงสร้างเนื้อหาวิชาต่างๆในระดับประถมศึกษาจะแบ่งตามมวลประสบการณ์ที่คิดว่าเป็นพื้นฐานความรู้ที่จำเป็นต่อเยาวชนอันได้แก่ กลุ่มทักษะที่เป็นเครื่องมือการเรียนรู้ กลุ่มสร้างเสริมประสบการณ์ชีวิต กลุ่มสร้างเสริมลักษณะนิสัย กลุ่มการทำงานพื้นฐานอาชีพ และกลุ่มประสบการณ์พิเศษ (กรมวิชาการ, 2535 ; ปานทอง กุลนาถศิริ, 2540)

ระเบียบกระทรวงศึกษาธิการ ว่าด้วยการประเมินผลการเรียนตามหลักสูตรประถมศึกษา พุทธศักราช 2521 (ฉบับปรับปรุง 2533) ได้กำหนดให้โรงเรียนมีหน้าที่รับผิดชอบการวัดและประเมินผล ทั้งการวัดและประเมินผลเพื่อปรับปรุงการเรียนการสอน และเพื่อตัดสินผลการเรียน โดยที่การวัดและประเมินผลเพื่อตัดสินผลการเรียนนั้นระเบียบกระทรวงศึกษาธิการ ว่าด้วยการประเมินผลการเรียนตามหลักสูตรประถมศึกษา พุทธศักราช 2521 (ฉบับปรับปรุง 2533) มีการกำหนดเกณฑ์หรือคะแนนจุดตัดไว้อย่างชัดเจน แต่การวัดและประเมินผลเพื่อปรับปรุงการเรียนการสอนด้วย

วิธีการให้ครูผู้สอนเป็นคนวัดและประเมินผลตามวัตถุประสงค์การเรียนรู้ว่าผู้เรียนมีความสามารถในเรื่องนั้นๆตามเกณฑ์หรือไม่ ยังไม่มีการกำหนดเกณฑ์หรือคะแนนจุดตัดที่บอกระดับความสามารถขั้นต่ำในการผ่านวัตถุประสงค์ที่แน่นอน (กรมวิชาการ, 2533) ดังนั้นการกำหนดเกณฑ์การผ่านจุดประสงค์ให้สอดคล้องกับความสามารถของผู้เรียน และลักษณะความยากง่ายของข้อสอบจึงเป็นหน้าที่ที่สำคัญประการหนึ่งของครูผู้สอน เพราะการกำหนดเกณฑ์ที่เหมาะสมจะช่วยให้การประเมินผลการเรียนรู้ของผู้เรียนเป็นไปอย่างถูกต้องและยุติธรรม อีกทั้งยังช่วยให้ครูนำผลการประเมินที่ได้มาแก้ไขข้อบกพร่องในการจัดกิจกรรมการเรียนการสอนให้มีประสิทธิภาพยิ่งขึ้น

เกลเซอร์ (Glaser, 1963 อ้างถึงใน Roid and Haladyna, 1982) เป็นคนแรกที่เสนอให้ใช้การทดสอบแบบอิงเกณฑ์ (Criterion - Referenced Testing) ในการสอนอย่างเป็นระบบโดยอธิบายหลักการของการทดสอบในแง่ของระดับเกณฑ์ และพฤติกรรมเฉพาะที่นักเรียนสามารถปฏิบัติได้โดยการสะท้อนออกมาในรูปของคะแนน สำหรับแนวความคิดเกี่ยวกับการวัดแบบอิงเกณฑ์ (Criterion - Referenced measurement) เกิดขึ้นจากการต้องการทราบว่าผลการสอบนั้นสามารถบอกได้หรือไม่ว่าผู้สอบมีความรอบรู้อยู่ในระดับใดของเนื้อหาทั้งหมดที่ทำกรวัด โดยการนำผลการวัดความสามารถของผู้สอบมาเปรียบเทียบกับเกณฑ์ที่กำหนดไว้ นอกจากนี้แล้วการวัดผลแบบอิงเกณฑ์นั้นจะต้องสอดคล้องกับวัตถุประสงค์ในการเรียนการสอนด้วย (สำเร็จ บุญเรืองรัตน์, 2527; Mehrens and Lemann, 1984; Bott, 1996) ในขณะที่การวัดผลแบบอิงกลุ่ม (Norm-Referenced Measurement) บอกให้เราทราบเพียงว่านักเรียนคนหนึ่งมีผลการทดสอบสูงหรือต่ำกว่านักเรียนคนอื่นๆเท่านั้น ไม่สามารถให้สารสนเทศที่จะตัดสินว่านักเรียนมีความรอบรู้หรือมีความสามารถเฉพาะงานตามจุดมุ่งหมายมากนักน้อยเพียงใด (Mehrens and Lemann, 1984; Wiersma and Jurs, 1990; Gellman, 1995; Bott, 1996) หรือกล่าวโดยสรุปว่าการวัดผลแบบอิงเกณฑ์มุ่งจะวัดความสามารถของผู้เรียนตามระดับการเรียนรู้ใน 2 ระดับ ดังนี้

1. วัดระดับความสามารถในการรอบรู้ (Mastery Level) คือ มุ่งจะวัดความสามารถในการเรียนรู้ตามเกณฑ์มาตรฐานที่กำหนดไว้
2. วัดระดับความสามารถในการพัฒนา (Development Level) คือ มุ่งจะวัดความสามารถในการเรียนรู้ของผู้เรียนที่สูงกว่าขั้นรอบรู้ การวัดในระดับนี้จะมุ่งชี้ถึงระดับความก้าวหน้าของผู้เรียนได้พัฒนาขึ้นมาจากการเรียนการสอนของครูผู้สอน

แบบสอบอิงเกณฑ์ (Criterion-Referenced Test) เป็นแบบสอบที่มุ่งวัดระดับการเรียนรู้ของผู้เรียนว่ามีความรู้ความสามารถอะไรบ้างและอยู่ในระดับใด การสร้างข้อสอบอิงเกณฑ์จึงต้องสร้างให้ครอบคลุมมวลความรู้ หรือทักษะสำคัญของการเรียนรู้ที่ต้องการให้เกิดขึ้นกับผู้เรียนอย่างมีลักษณะเฉพาะ (บุญธรรม กิจบริตาภิสุทธิ์, 2535; ศิริชัย กาญจนวาสี, 2535) และเกลเซอร์ (Glaser, 1963 อ้างถึงใน เขาวดี วิบูลย์ศรี, 2539) กล่าวไว้ว่า แบบสอบอิงเกณฑ์จะต้องมีเกณฑ์เพื่อวินิจฉัยว่า

นักเรียนคนใดเรียนรู้ครบถ้วนตามวัตถุประสงค์หรือไม่ และเกณฑ์หรือคะแนนจุดตัด (Cut-off score) ที่ใช้โดยทั่วไป ควรมีคุณสมบัติ 2 ประการ คือ

1. สามารถบรรยายถึงวัตถุประสงค์ของการเรียนการสอนได้อย่างชัดเจน
2. สามารถบ่งชี้ถึงประสิทธิภาพของพฤติกรรมอันเป็นที่ยอมรับได้ หรือบอกถึงระดับของการปฏิบัติที่เป็นมาตรฐาน ในขณะที่เดียวกันผลจากการทดสอบแบบอิงเกณฑ์จะต้องสามารถนำไปตีความลักษณะงานเฉพาะที่บุคคลกระทำได้อย่างอิสระโดยไม่จำเป็นต้องอ้างอิงถึงกลุ่มปกติวิสัย หรือกล่าวได้ว่าคะแนนที่ได้จากแบบสอบอิงเกณฑ์มักจะตีความหมายในเชิงสัมบูรณ์ (Absolute) มากกว่าที่จะตีความหมายในลักษณะของมาตรฐานเชิงสัมพัทธ์ (Relative standard) ดังที่ปฏิบัติกันในการตีความหมายของคะแนนที่ได้จากแบบสอบอิงกลุ่ม

จากแนวคิดเกี่ยวกับการวัดและลักษณะของแบบสอบอิงเกณฑ์ดังกล่าว จะเห็นได้ว่าการกำหนดมาตรฐาน (Standard Setting) หรือคะแนนจุดตัด (Cut-off score) เพื่อใช้เป็นเกณฑ์บอกถึงระดับความสามารถของผู้เรียนว่ามีความรอบรู้ในลักษณะงานเฉพาะนั้นๆหรือไม่ ถือเป็นคุณสมบัติที่สำคัญของการวัดผลแบบอิงเกณฑ์ แต่การกำหนดคะแนนจุดตัดยังเป็นที่โต้แย้งกันมากระหว่างนักวัดผลว่าเป็นวิธีการที่ไม่มีเหตุผลที่ชัดเจนตลอดจนมีความเป็นอัตนัยสูง ดังที่ กลาส (Glass, 1978) กล่าวไว้ว่าการกำหนดมาตรฐานหรือคะแนนจุดตัดเป็นการกำหนดเกณฑ์ขึ้นเองตามอำเภอใจ เช่นการกำหนดเกณฑ์ให้ 80% เป็นคะแนนที่ผ่านนั้นไม่สามารถอธิบายโดยอาศัยเหตุผลตามหลักการทางด้านการวัดผลได้ชัดเจน ฉะนั้นการถือเกณฑ์ตามที่กำหนดให้ผ่านจึงไม่มีมาตรฐานที่แน่นอนและการให้คะแนนขึ้นอยู่กับความเห็นส่วนบุคคลเป็นหลัก เพราะการให้คะแนนนั้นเน้นที่ตัวบุคคลไม่มีการเปรียบเทียบกับกลุ่ม

ตามปกติแล้วเกณฑ์มาตรฐานหรือคะแนนจุดตัดที่ใช้ในการวัดระดับการเรียนรู้มักจะขาดความน่าเชื่อถือถ้าไม่สามารถหาเหตุผลตามหลักตรรกศาสตร์มาสนับสนุน หรือไม่ก็เป็นไปตามการสังเกตอย่างมีระบบทางวิทยาศาสตร์ ดังตัวอย่างข้อสอบต่อไปนี้ (เยาวดี วิบูลย์ศรี, 2539)

ข้อสอบข้อที่ 1 "ลีด้าสามารถรับความร้อนได้ดีกว่าลิวา" (ถูกหรือผิด)

ข้อสอบข้อที่ 2 "ลูกบอลลูน 2 ลูกมีขนาดเท่ากัน บรรจุก๊าซเท่ากันแต่ต่างกันที่บอลลูนใบหนึ่งลิวา ส่วนอีกใบหนึ่งลีด้า ถ้านำบอลลูนทั้ง 2 ลูกไปปรับแสงเท่ากันผลจะเป็นอย่างไร"

- ก. ทั้ง 2 ลูกมีขนาดเล็กลง
- ข. ทั้ง 2 ลูกมีขนาดเท่ากัน
- ค. ลูกลีด้าจะมีขนาดโตกว่าลูกลิวา
- ง. ลูกลิวาจะมีขนาดโตกว่าลูกลีด้า

จากตัวอย่างข้อสอบข้างต้นชี้ให้เห็นว่า เราไม่สามารถที่จะตัดสินตามเกณฑ์มาตรฐานหรือคะแนนจุดตัดเดียวกันในคำถามทั้ง 2 ข้อได้เนื่องจากระดับความยากง่ายของข้อสอบไม่เท่ากัน ดังจะเห็นได้ว่า คำถามในข้อสอบข้อที่ 2 ต้องอาศัยความรู้ความเข้าใจและสามารถที่จะนำความรู้ไปประยุกต์ใช้ในสถานการณ์ใหม่ ส่วนคำถามข้อที่ 1 เพียงแต่มีความรู้ก็สามารถตอบได้ ฉะนั้นถ้าถือเกณฑ์ 80% เป็นคะแนนผ่านก็อาจจะขาดมาตรฐานในการวัดผล เพราะถึงแม้ว่าผู้สอบได้คะแนนในระดับเดียวกันตามเกณฑ์ดังกล่าวก็ไม่จำเป็นจะต้องมีความรู้เท่ากัน ถ้าหากข้อสอบนั้นมีระดับความยากง่ายที่แตกต่างกัน

ด้วยเหตุผลดังกล่าว จึงทำให้การกำหนดมาตรฐาน (Standard Setting) หรือคะแนนจุดตัด (Cut-off score) จึงเป็นประเด็นที่นักวัดผลให้ความสนใจและพยายามศึกษาวิธีการที่ทำให้การกำหนดคะแนนจุดตัดเป็นไปตามหลักการและมีเหตุผลตามหลักการของการวัดผล การกำหนดคะแนนจุดตัดของแบบสอบอิงเกณฑ์นั้น ได้มีผู้เสนอแนวทางไว้หลายวิธีด้วยกัน สามารถสรุปตามวิธีดำเนินการแบบกว้างๆ ได้ 3 ประเภทดังนี้ (Hambleton, 1980 ; Suen, 1990)

1. คะแนนจุดตัดจากการใช้ดุลยพินิจของผู้เชี่ยวชาญในเนื้อหา หรือผู้เกี่ยวข้องในเนื้อหา (Judgmental Methods) การกำหนดมาตรฐานวิธีนี้จะได้คะแนนจุดตัดที่คงที่ไม่แปรเปลี่ยนไปตามคะแนนของผู้สอบซึ่งเป็นการกำหนดจุดตัดแบบสัมบูรณ์ (Absolute) ได้แก่ วิธีของนิเดลสกี (Nedelsky's Method) วิธีของอีเบล (Ebel's Method) วิธีแองกอฟ (Angoff's Method) และวิธีที่พัฒนาจากวิธีแองกอฟ (Modified Angoff's Method)

2. การกำหนดคะแนนจุดตัดโดยอาศัยข้อมูลจากการสอบของกลุ่มผู้สอบ (Empirical Model) ผลการกำหนดเกณฑ์ด้วยวิธีนี้มีค่าแปรเปลี่ยนหรือสัมพันธ์กับคะแนนจากการสอบของกลุ่มผู้สอบ ได้แก่ วิธีของลิฟวิงสตัน (Livingston's Method) และวิธีของไคววอล (Kriewal's Method)

3. การกำหนดคะแนนจุดตัดด้วยการตัดสินของผู้เชี่ยวชาญรวมกับการใช้ข้อมูลจากการสอบของกลุ่มผู้สอบ (Combination Model) ได้แก่ วิธีของลิฟวิงสตันและซายกี (Livingston and Zieky's Method) วิธีหาจุดสมดุลระหว่างเกณฑ์สัมบูรณ์กับเกณฑ์สัมพัทธ์ของบูค (Beuk's Method) วิธีตัดสินโดยอาศัยสารสนเทศประกอบ (Informed Judgment) วิธีของเบอร์ก (Berk's Method) และวิธีของบลอค (Block's Method)

กลาส (Glass, 1978) เสนอแนวคิดไว้ว่า วิธีการกำหนดคะแนนจุดตัดที่อาศัยข้อมูลจากการสอบของกลุ่มผู้สอบโดยการพิจารณาความสามารถของผู้สอบคนอื่นๆ แล้วกำหนดเป็นจุดตัดนั้น เป็นวิธีที่ไม่แตกต่างจากการประเมินผลแบบอิงกลุ่ม สำหรับบางวิธีที่อาศัยข้อมูลจากการสอบนำไปสัมพันธ์กับเกณฑ์ภายนอกอื่นๆ เช่น วิธีของบลอค (Block's Method) และวิธีของเบอร์ก (Berk's Method)

จะใช้ได้ผลดีก็ต่อเมื่อตัวเกณฑ์ภายนอกกับคะแนนจากแบบสอบถามมีความสัมพันธ์กันแบบสัมบูรณ์ ส่วนวิธีการกำหนดเกณฑ์ที่ใช้ดุลยพินิจของผู้เชี่ยวชาญนั้น กลาส มีความเห็นที่สอดคล้องกับ แวน เดอร์ ลินเดน (Van der Linden, 1982) ว่ามีปัญหาที่สำคัญคือ ความสามารถของผู้เชี่ยวชาญในการกำหนดความสามารถขั้นต่ำสุดของผู้เรียนและความสอดคล้องหรือความเที่ยงในการตัดสินของผู้เชี่ยวชาญ

จากปัญหาของการกำหนดคะแนนจุดตัดในแต่ละวิธีดังกล่าวข้างต้น จึงทำให้มีผู้สนใจศึกษาเพื่อหาวิธีแก้ปัญหในการกำหนดคะแนนจุดตัดเป็นจำนวนมาก เช่น อัจฉริยา ปราบอรทัย (2531) ได้ศึกษาเปรียบเทียบคะแนนจุดตัด และความคงที่หรือความสอดคล้องในการตัดสินคะแนนจุดตัด (Indices of the consistency) ด้วยวิธีของแองกอฟเมื่อให้และไม่ให้สารสนเทศแก่ผู้เชี่ยวชาญ ประกอบการตัดสิน โดยการประยุกต์ตามทฤษฎีการตอบสนองข้อสอบ พบว่าการให้และไม่ให้สารสนเทศแก่ผู้เชี่ยวชาญมีผลทำให้คะแนนจุดตัดแตกต่างกันอย่างมีนัยสำคัญทางสถิติ แต่ค่าความคงที่ของการตัดสินใจไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ เฟรดแมนและโฮ (Friedman and Ho, 1990) ให้ผู้เชี่ยวชาญจำนวน 11 คน จาก 11 รัฐของอเมริกาตัดสินคะแนนจุดตัด โดยใช้ข้อสอบจำนวน 65 ข้อ หลังจากนั้นนำผลการตัดสินที่ได้มาให้ผู้เชี่ยวชาญทั้งหมดร่วมกันอภิปราย หลังจากนั้นจึงให้ผู้เชี่ยวชาญตัดสินคะแนนจุดตัดของข้อสอบชุดเดิมอีกครั้ง พบว่า สามารถเพิ่มความสอดคล้องในการตัดสินระหว่างกลุ่มผู้เชี่ยวชาญ (Interjudges Consistency) และความสอดคล้องในการตัดสินภายในของผู้เชี่ยวชาญ (Intrajudges Consistency) ด้วย

เพลคและอิมพารา (Plake and Impara, 1996) ศึกษาความสอดคล้องภายในของผู้ตัดสิน (Intrajudge consistency) ของการกำหนดคะแนนจุดตัดด้วยวิธีของแองกอฟ โดยให้ผู้เชี่ยวชาญกำหนดสัดส่วนในการตอบข้อสอบถูกของผู้เรียนที่มีระดับความสามารถต่ำสุดที่ยอมรับได้ โดยใช้ข้อสอบ 24 ข้อจำนวน 2 ครั้ง ผลสรุปชี้ให้เห็นว่ามีความสอดคล้องภายในของผู้เชี่ยวชาญสูงในกรณีความแตกต่างของค่าเฉลี่ยสัมบูรณ์ในการประมาณค่าการปฏิบัติข้อสอบ และคะแนนจุดตัด แสดงให้เห็นว่า คุณลักษณะของการศึกษาเกณฑ์ และคุณสมบัติของแบบสอบถามมีผลต่อระดับความสอดคล้องภายในของผู้ตัดสิน

อิมพาราและเพลค (Impara and Plake, 1997) เปรียบเทียบวิธีการกำหนดมาตรฐานระหว่างวิธีของแองกอฟแบบดั้งเดิม (The traditional Angoff's method) และวิธีของอิมพาราและเพลค (Impara and Plak's Method) หรือที่เรียกว่าวิธีตัดสินว่าได้หรือตก (The yes-no method) พบว่า วิธีกำหนดคะแนนทั้งสองให้คะแนนจุดตัดที่ใกล้เคียงกัน และผู้เชี่ยวชาญเห็นว่าวิธีของอิมพาราและเพลค มีความสะดวกในการปฏิบัติมากกว่าวิธีของแองกอฟแบบดั้งเดิม

เพลคและเคน (Plake and Kane, 1991) เปรียบเทียบวิธีการกำหนดคะแนนจุดตัดหรือระดับการผ่านต่ำสุดที่ยอมรับได้จากการประมาณข้อสอบแต่ละข้อของผู้เชี่ยวชาญ วิธีที่ใช้คือ วิธีที่ 1

ให้นำหน้าระดับการผ่านต่ำสุดที่ยอมรับได้สำหรับข้อสอบตามระดับความสามารถของผู้สอบ วิธีที่ 2 ให้นำหน้าการประมาณระดับการผ่านต่ำสุดที่ยอมรับได้ของข้อสอบเท่ากัน ซึ่งเป็นวิธีกำหนด ระดับการผ่านต่ำสุดที่ยอมรับได้การประมาณของผู้เชี่ยวชาญแบบดั้งเดิม และวิธีที่ 3 ผลผสมระหว่างการทำนายระดับการผ่านต่ำสุดที่ยอมรับได้จากผู้เชี่ยวชาญและสัดส่วนของความสามารถต่ำสุดของผู้สอบที่ตอบข้อสอบแต่ละข้อได้ถูก พบว่าทั้งสามวิธีประมาณค่าได้ถูกต้องตามระดับความสามารถ โดยทั้งหมดมีความสอดคล้องในการตัดสินสูงสุดเมื่อระดับความสามารถเข้าคู่กับความยากโดยเฉลี่ยของข้อสอบ และจำนวนผู้เชี่ยวชาญ 10 คนจะมีความสอดคล้องในการตัดสินสูงกว่าผู้เชี่ยวชาญ 5 คน แต่ข้อสอบจำนวน 50 ข้อ และ 25 ข้อมีความสอดคล้องในการตัดสินไม่แตกต่างกัน

จากการศึกษาแนวคิดและงานวิจัยที่เกี่ยวข้องกับการคะแนนจุดตัดดังกล่าว จะเห็นว่าวิธีคะแนนจุดตัดแต่ละวิธีต่างมีข้อดีและข้อเสียที่แตกต่างกันไป และพอปแฮม (Popham, 1978 อ้างถึงใน Hambleton, 1980) กล่าวว่า ถึงแม้ Glass จะเห็นว่า การกำหนดคะแนนจุดตัดเป็นการตัดสินที่ตามอำเภอใจ (arbitrary) แต่ก็เป็นที่ยอมรับกันว่ามาตรฐานในการปฏิบัติจะต้องมีการตัดสิน และ Webster's Dictionary ได้ให้ความหมายของคำว่า ตามอำเภอใจ (arbitrary) ไว้ 2 ลักษณะคือ ความหมายแรกเป็นด้านบวก หมายถึง คำคุณศัพท์ที่แสดงถึงตัวเลือกหรือการวินิจฉัยเคราะห์ที่สามารถตัดสินได้โดยผู้ตัดสิน (Judge) ความหมายที่สองคือ คำคุณศัพท์ที่แสดงถึงการทำตามอำเภอใจ เช่นการเลือกอย่างสุ่มและตามอำเภอใจ ซึ่งถ้ามองตามความหมายที่หนึ่งจะสะท้อนให้เห็นถึงความถูกต้องของการพยายามกำหนดคะแนนจุดตัด ดังนั้น Popham จึงเห็นว่า การกำหนดคะแนนจุดตัดไม่สามารถหลีกเลี่ยงการใช้ดุลยพินิจตัดสินได้ และการกล่าวว่าการใช้ดุลยพินิจตัดสินเป็นไปโดยปราศจากเหตุผลก็เป็นเรื่องที่ไม่ถูกต้องทั้งหมดเพราะถึงแม้ว่าวิธีกำหนด คะแนนจุดตัดโดยใช้ดุลยพินิจของผู้เชี่ยวชาญจะมีข้อจำกัด แต่จากการศึกษาพบว่าวิธีกำหนดคะแนนจุดตัดด้วยวิธีดังกล่าวเป็นพื้นฐานสำคัญที่นำไปสู่การกำหนดมาตรฐานด้วยวิธีอื่น ๆ อีกทั้งยังง่ายต่อการปฏิบัติด้วย จึงทำให้มีผู้สนใจศึกษาเพื่อแก้ไขข้อบกพร่องของวิธีดังกล่าว เช่น แพทริเซีย (Patricia, 1991) พบว่าการกำหนดคะแนนจุดตัดโดยใช้ดุลยพินิจของผู้เชี่ยวชาญขึ้นอยู่กับคุณภาพของผู้ตัดสินหรือผู้เชี่ยวชาญ และข้อมูลสารสนเทศประกอบารตัดสินที่เพียงพอ

ด้วยเหตุผลดังกล่าวมาข้างต้นจึงทำให้ผู้วิจัยสนใจที่จะศึกษาการตัดสินคะแนนจุดตัดของผู้เชี่ยวชาญ โดยการนำสารสนเทศเกี่ยวกับคุณภาพของข้อสอบในเรื่องของค่าความยาก อำนาจจำแนก และประสิทธิภาพตัวกลางที่วิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิมมาใช้ประกอบในการตัดสินคะแนนจุดตัดว่าจะช่วยเพิ่มคุณภาพในเรื่องของความสอดคล้อง หรือความเที่ยงในการตัดสินของผู้เชี่ยวชาญตามที่ มีนักการศึกษาได้เสนอไว้หรือไม่ นอกจากนี้ผู้วิจัยยังสนใจที่จะศึกษาว่าเมื่อใช้กลุ่มผู้เชี่ยวชาญเป็นครูผู้สอนซึ่งมีความคุ้นเคยใกล้ชิด และรู้จักผู้เรียนเป็นอย่างดีจะตัดสินคะแนนจุดตัดได้ถูกต้องกับคะแนน

จากการสอบของผู้เรียนเพียงใด และจากการศึกษาของ เคน (Kane, 1987) ที่ได้เปรียบเทียบวิธีวิเคราะห์ผลการตัดสินคะแนนจุดตัดด้วยวิธีของแองกอฟ พบว่า วิธีการวิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิมไม่ใช่วิธีที่ดีที่สุด และได้สนับสนุนให้ใช้วิธีวิเคราะห์ตามทฤษฎีการสรุปอ้างอิง (Generalizability Theory) ดังนั้นผู้วิจัยจึงสนใจที่จะศึกษาความสอดคล้องหรือความเที่ยงในการตัดสินใจโดยการประยุกต์วิธีวิเคราะห์ตามทฤษฎีการสรุปอ้างอิงทางการทดสอบ เพราะทฤษฎีการสรุปอ้างอิงเป็นทฤษฎีที่ขยายแนวคิดของการประมาณค่าความเที่ยงของการทดสอบตามทฤษฎีการทดสอบแบบดั้งเดิมให้ครอบคลุมเงื่อนไขของการวัด และแหล่งความคลาดเคลื่อนต่างๆโดยเครื่องมือที่ใช้ไม่จำเป็นต้องตั้งอยู่บนข้อตกลงเบื้องต้นของแบบสอบคู่ขนาน โดยทฤษฎีการสรุปอ้างอิงอาศัยหลักการจำแนกความคลาดเคลื่อนของการวัดออกเป็นความคลาดเคลื่อนอันเกิดจากหลายแหล่ง เพื่อประมาณค่าสัมประสิทธิ์การสรุปอ้างอิงหรือสัมประสิทธิ์ความเที่ยงทั่วไปภายใต้เงื่อนไขของการวัดแบบต่างๆ (ศิริชัย กาญจนวาสี, 2535) โดยการวิจัยครั้งนี้มีวัตถุประสงค์ของการวิจัยดังนี้

วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาคะแนนจุดตัดที่ได้จากการตัดสินของผู้เชี่ยวชาญที่มีสารสนเทศเกี่ยวกับคุณภาพของข้อสอบในเรื่องของค่าความยาก อำนาจจำแนก และประสิทธิภาพตัวลงที่วิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิมมาใช้ประกอบในการตัดสินคะแนนจุดตัดด้วยวิธีของแองกอฟ วิธีของนิตลสกี และวิธีของอิมพาราและเพลค
2. เพื่อศึกษาความสัมพันธ์ระหว่างการตัดสินคะแนนจุดตัดของผู้เชี่ยวชาญที่มีสารสนเทศเกี่ยวกับคุณภาพของข้อสอบในเรื่องของค่าความยาก อำนาจจำแนก และประสิทธิภาพตัวลงที่วิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิมมาใช้ประกอบในการตัดสินคะแนนจุดตัดด้วยวิธีของแองกอฟ วิธีของนิตลสกี และวิธีของอิมพาราและเพลคกับคะแนนความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องของผู้เรียน
3. เพื่อศึกษาความสัมพันธ์ระหว่างการตัดสินคะแนนจุดตัดของผู้เชี่ยวชาญที่มีสารสนเทศเกี่ยวกับคุณภาพของข้อสอบในเรื่องของค่าความยาก อำนาจจำแนก และประสิทธิภาพตัวลงที่วิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิมมาใช้ประกอบในการตัดสินคะแนนจุดตัดด้วยวิธีของแองกอฟ วิธีของนิตลสกี และวิธีของอิมพาราและเพลคกับคะแนนความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องของผู้เรียนในกลุ่มที่รอบรู้และไม่รอบรู้จริงตามทฤษฎีการตัดสินใจ (Decision Theoretic Approaches)

4. เพื่อเปรียบเทียบความสอดคล้องในการตัดสินคะแนนจุดตัดของผู้เชี่ยวชาญที่มีสารสนเทศเกี่ยวกับคุณภาพของข้อสอบในเรื่องของค่าความยาก อำนาจจำแนก และประสิทธิภาพตัวลงที่วิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิมมาใช้ประกอบในการตัดสินคะแนนจุดตัด ระหว่างวิธีของแองกอฟ วิธีของนิเดิลสกี และวิธีของอิมพาราและเพลค

สมมุติฐานในการวิจัย

จากงานวิจัยของ ครอสและคณะ (Cross, et al., 1984) ศึกษาความสอดคล้องในการกำหนดคะแนนจุดตัด (Indices of the consistency) ระหว่างวิธีของแองกอฟ วิธีของนิเดิลสกี และวิธีของเจเกอร์ โดยวิเคราะห์ความสอดคล้องในการกำหนดคะแนนจุดตัดด้วยการประยุกต์ตามทฤษฎีการสรุปอ้างอิง (Generalizability Theory) พบว่าวิธีของแองกอฟมีความสอดคล้องในการกำหนดคะแนนจุดตัดสูงที่สุด ส่วนวิธีของนิเดิลสกีมีความสอดคล้องในการกำหนดคะแนนจุดตัดต่ำสุด เคน (Kane, 1987) ที่ได้เปรียบเทียบวิธีวิเคราะห์ผลการตัดสินคะแนนจุดตัดด้วยวิธีของแองกอฟ พบว่าวิธีการวิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิมไม่ใช่วิธีที่ดีที่สุด และได้สนับสนุนให้ใช้วิธีวิเคราะห์ตามทฤษฎีการสรุปอ้างอิง (Generalizability Theory) และอัจฉริยา ปราบอริพ่าย (2531) ศึกษาเปรียบเทียบการกำหนดคะแนนจุดตัดวิธีแองกอฟเมื่อให้และไม่ให้สารสนเทศแก่ผู้เชี่ยวชาญเพื่อประกอบการตัดสิน พบว่าการให้และไม่ให้สารสนเทศแก่ผู้เชี่ยวชาญมีผลทำให้คะแนนจุดตัดแตกต่างกันอย่างมีนัยสำคัญทางสถิติและค่าความสอดคล้องหรือความคงที่ในการตัดสิน (Indices of the consistency) ไม่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ แพทริเซีย (Patricia, 1991) พบว่าการกำหนดคะแนนจุดตัดโดยใช้ดุลยพินิจในการตัดสินขึ้นอยู่กับคุณภาพของผู้ตัดสิน และข้อมูลสารสนเทศประกอบการตัดสินที่เพียงพอ และนอร์ซินีและคณะ (Norcini, et al, 1991) ศึกษาผลของจำนวนผู้เชี่ยวชาญและจำนวนข้อสอบรวมที่ใช้ในการตัดสินคะแนนจุดตัดโดยผู้เชี่ยวชาญ พบว่า ผู้เชี่ยวชาญ 5 คนและข้อสอบรวม 25 ข้อ จะทำให้ความคลาดเคลื่อนลดลง ผู้วิจัยจึงกำหนดสมมุติฐานในการวิจัยดังต่อไปนี้

ความสอดคล้องในการตัดสินคะแนนจุดตัดของผู้เชี่ยวชาญที่มีสารสนเทศเกี่ยวกับคุณภาพของข้อสอบในเรื่องของค่าความยาก อำนาจจำแนก และประสิทธิภาพตัวลงที่วิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิมมาใช้ประกอบในการตัดสินคะแนนจุดตัดระหว่างวิธีของแองกอฟ วิธีของนิเดิลสกี และวิธีของอิมพาราและเพลค น่าจะแตกต่างกัน

ขอบเขตของการวิจัย

1. ประชากรที่ใช้ในการวิจัย แบ่งเป็น 2 กลุ่ม ดังนี้
 - 1.1 ประชากรผู้เชี่ยวชาญ ประกอบด้วย ครูผู้สอนวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2541 ในโรงเรียนสังกัดสำนักงานการประถมศึกษาจังหวัดเลย โดยที่ผู้เชี่ยวชาญจะต้องมีคุณสมบัติดังนี้
 1. เป็นครูผู้สอนที่สอนวิชาคณิตศาสตร์แก่นักเรียนกลุ่มตัวอย่างตั้งแต่ภาคเรียนที่ 1 ปีการศึกษา 2541 จนถึงปัจจุบัน
 2. มีประสบการณ์ในการสอนวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 ตั้งแต่ 5 ปีขึ้นไป หรือเคยผ่านการอบรมเกี่ยวกับการสอนวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6
 - 1.2 ประชากรนักเรียน ประกอบด้วยนักเรียนชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2541 ในโรงเรียนประถมศึกษา สังกัดสำนักงานการประถมศึกษาจังหวัดเลย จำนวน 8,771 คน

2. ตัวแปรที่ศึกษา

ตัวแปรต้น คือ วิธีการมาตรฐานหรือคะแนนจุดตัด ได้แก่ วิธีของแองกอฟ วิธีของนิเดิลสกี และวิธีของอิมพาราและเพลค

ตัวแปรตาม คือ คะแนนจุดตัด ความสอดคล้องในการตัดสินคะแนนจุดตัดของผู้เชี่ยวชาญ

3. ทฤษฎีที่ใช้ในการวิจัย ได้แก่ ทฤษฎีการสรุปอ้างอิง

4. ผู้วิจัยดำเนินการวิจัยด้วยการเก็บรวบรวมข้อมูลจากกลุ่มตัวอย่างจำนวน 2 ครั้ง เพื่อต้องการให้ได้ข้อมูลที่สามารถยืนยันถึงความสอดคล้องในการตัดสินคะแนนจุดตัด โดยที่แต่ละครั้งมีระยะเวลาห่างกันประมาณ 2 สัปดาห์ สำหรับการเก็บรวบรวมข้อมูลจากกลุ่มตัวอย่างผู้เชี่ยวชาญ ผู้วิจัยได้มอบหมายให้ผู้เชี่ยวชาญแต่ละคนตัดสินข้อสอบเป็นรายข้อเพื่อนำมากำหนดเป็นคะแนนจุดตัดของแบบสอบทั้งหมด หลังจากนั้นจึงนำคะแนนจุดตัดมาวิเคราะห์หาค่าความสอดคล้องโดยการประยุกต์ตามวิธีการหาค่าสัมประสิทธิ์ความเที่ยงด้วยทฤษฎีการสรุปอ้างอิง

คำจำกัดความที่ใช้ในการวิจัย

1. ผู้เชี่ยวชาญ หมายถึง ครูผู้สอนวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 ในโรงเรียนประถมศึกษา สังกัดสำนักงานการประถมศึกษาจังหวัดเลย โดยเป็นครูผู้สอนที่สอนวิชาคณิตศาสตร์แก่นักเรียนกลุ่มตัวอย่างตั้งแต่ภาคเรียนที่ 1 ปีการศึกษา 2541 จนถึงปัจจุบัน และมีประสบการณ์

ในการสอนวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 ตั้งแต่ 5 ปีขึ้นไป หรือเคยผ่านการอบรมเกี่ยวกับการสอนวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6

2. คะแนนจุดตัด หมายถึง คะแนนที่ใช้เป็นเกณฑ์ผ่านชั้นต่ำสุดที่จะตัดสินความสามารถของผู้เรียนว่าเป็นผู้รอบรู้หรือไม่รอบรู้ ผ่านหรือตกในขอบเขตเนื้อหาที่ต้องการวัด และมีการกำหนดคะแนนจุดตัดจากการให้ผู้เชี่ยวชาญพิจารณาข้อสอบแต่ละข้อพร้อมทั้งสารสนเทศเกี่ยวกับคุณภาพของข้อสอบในเรื่องของค่าความยาก อำนาจจำแนก และประสิทธิภาพตัวลงที่วิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิมมาใช้ประกอบในการตัดสินคะแนนจุดตัด โดยที่วิธีของแองกอฟจะกำหนดให้ผู้เชี่ยวชาญพิจารณาข้อสอบแต่ละข้อพร้อมทั้งสารสนเทศเกี่ยวกับคุณภาพของข้อสอบแล้วตัดสินประมาณค่าร้อยละหรือความน่าจะเป็นของกลุ่มผู้สอบที่มีผลการเรียนในระดับคาบเส้นหรือมีผลการเรียนเป็น 0 และ 1 จะสามารถทำข้อสอบนั้นได้ถูกต้อง ซึ่งในการวิจัยครั้งนี้ได้กำหนดให้ผู้เชี่ยวชาญตัดสินโดยการประมาณค่าร้อยละของกลุ่มผู้สอบที่มีผลการเรียนในระดับคาบเส้นหรือมีผลการเรียนเป็น 0 และ 1 จะสามารถทำข้อสอบนั้นได้ถูกต้อง หลังจากนั้นจึงนำค่าร้อยละที่ได้มาเปลี่ยนเป็นค่าความน่าจะเป็นในการตอบข้อสอบถูก วิธีของนิคดลสก็กำหนดให้ผู้เชี่ยวชาญพิจารณาข้อสอบแต่ละข้อพร้อมทั้งสารสนเทศเกี่ยวกับคุณภาพของข้อสอบโดยพิจารณาตัวเลือกว่า มีตัวเลือกใดบ้างที่ผู้สอบที่มีผลการเรียนในระดับคาบเส้นหรือมีผลการเรียนเป็น 0 และ 1 บอกได้ว่าผิด และวิธีของอิมพาราและเพลลจะกำหนดให้ผู้เชี่ยวชาญพิจารณาข้อสอบแต่ละข้อพร้อมทั้งสารสนเทศเกี่ยวกับคุณภาพของข้อสอบ แล้วตัดสินว่ากลุ่มผู้สอบที่มีผลการเรียนในระดับคาบเส้นหรือมีผลการเรียนเป็น 0 และ 1 สามารถตอบข้อคำถามแต่ละข้อนั้นได้ถูกหรือผิด

3. สารสนเทศประกอบการตัดสิน หมายถึง สารสนเทศเกี่ยวกับคุณภาพของข้อสอบที่วิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory) ได้แก่ ค่าความยาก ค่าอำนาจจำแนก หรือดัชนีความไว และประสิทธิภาพของตัวลง

4. กลุ่มผู้เรียนที่รอบรู้และไม่รอบรู้จริงตามทฤษฎีการตัดสินใจ (Decision Theoretic Approaches) หมายถึง กลุ่มผู้เรียนที่ได้คะแนนสูงกว่าหรือเท่ากับคะแนนจุดตัดตามวิธีการใช้ทฤษฎีการตัดสินใจ (Decision Theoretic Approaches) และผ่านเกณฑ์ภายนอก (ผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ ภาคเรียนที่ 1 ปีการศึกษา 2541) กับกลุ่มผู้เรียนที่ได้คะแนนต่ำกว่าคะแนนจุดตัดตามวิธีการใช้ทฤษฎีการตัดสินใจ (Decision Theoretic Approaches) และไม่ผ่านเกณฑ์ภายนอก (Suen, 1990)

5. ทฤษฎีการสุบอ้างอิงทางการทดสอบ หมายถึง ทฤษฎีการวัดอาศัยหลักการจำแนกความคลาดเคลื่อนของการวัดออกเป็นความคลาดเคลื่อนอันเกิดจากหลายแหล่ง เพื่อประมาณค่าสัมประสิทธิ์การสุบอ้างอิงหรือสัมประสิทธิ์ความเที่ยงทั่วไปภายใต้เงื่อนไขของการวัดแบบต่างๆ

6. เอกภาพ หมายถึง เงื่อนไขที่เป็นไปได้ทั้งหมดในการวัด ได้แก่ จำนวนผู้ตัดสินทั้งหมด และวิธีกำหนดมาตรฐานหรือคะแนนจุดตัดทุกวิธี
7. ฟาเซท หมายถึง องค์ประกอบหรือลักษณะร่วมกันของกลุ่มเงื่อนไขการวัดชนิดเดียวกัน ในการวิจัยครั้งนี้ มี 2 ฟาเซท ได้แก่ จำนวนผู้ตัดสิน และวิธีกำหนดคะแนนจุดตัด
8. เงื่อนไขการวัด หมายถึง สภาพการณ์ของการตัดสิน สำหรับการวิจัยนี้ได้แก่ ผู้ตัดสินแต่ละคน และวิธีกำหนดคะแนนจุดตัดที่สุ่มมาใช้
9. เอกภาพของการสังเกตที่ยอมรับได้ หมายถึง กลุ่มเงื่อนไขการวัดที่เป็นไปได้ทั้งหมด ได้แก่ ครูผู้สอนวิชาคณิตศาสตร์ชั้นประถมศึกษาปีที่ 6 ทุกคน ข้อสอบทุกข้อ และวิธีกำหนดคะแนนจุดตัดทุกวิธี
10. เอกภาพของการสรุปอ้างอิง หมายถึง กลุ่มเงื่อนไขของฟาเซทที่ผู้วิจัยต้องการสรุปอ้างอิงผลการวัด สำหรับการวิจัยนี้ ได้แก่ ผู้ตัดสินและวิธีกำหนดคะแนนจุดตัด
11. กลุ่มผู้เรียนที่มีความสามารถต่ำสุดที่ยอมรับได้ หมายถึง กลุ่มผู้เรียนที่อยู่ในระดับคาบเส้นหรือกลุ่มผู้เรียนที่มีผลการเรียนอยู่ในระดับ 0 และ 1
12. ความสอดคล้องในการตัดสิน (Indices of consistency) หมายถึง ความเที่ยงหรือความคงที่ในการตัดสินคะแนนจุดตัดระหว่างผู้เชี่ยวชาญ มีค่าเท่ากับสัดส่วนระหว่างความแปรปรวนของคะแนนเอกภาพ กับความแปรปรวนของค่าคาดหมายของคะแนนที่สังเกตได้ตามทฤษฎีการสรุปอ้างอิง
13. ความตรงในการตัดสินของผู้เชี่ยวชาญ หมายถึง ความสามารถของผู้เชี่ยวชาญในการตัดสินคะแนนจุดตัดได้ถูกต้องกับคะแนนความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องของผู้เรียน ซึ่งพิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างการตัดสินคะแนนจุดตัดของผู้เชี่ยวชาญกับคะแนนความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องของผู้เรียน

ประโยชน์ที่คาดว่าจะได้รับ

1. เป็นการตรวจสอบการกำหนดคะแนนจุดตัดที่ได้จากการตัดสินของผู้เชี่ยวชาญ ซึ่งเป็นครูผู้สอนที่มีความคุ้นเคยและใกล้ชิดกับนักเรียนว่ามีความถูกต้องเพียงใด
2. เป็นแนวทางในการสร้างทางเลือกวิธีการกำหนดคะแนนจุดตัดมีหลักการ และมีความสะดวกในทางปฏิบัติโดยไม่ต้องอาศัยวิธีการคำนวณที่ยุ่งยากสำหรับผู้สนใจ ตลอดจนช่วยเพิ่มประสิทธิภาพในการวัดผลแบบอิงเกณฑ์