บทที่ 2

วรรณคดีที่เกี่ยวข้อง

ผู้วิจัยได้ศึกษาและรวบรวมเอกสารที่เกี่ยวข้องจากหนังสือ เอกสาร และวารสารต่างๆโดยแบ่งการนำเสนอออกเป็น 3 ตอนดังต่อไปนี้

ตอนที่ 1 แนวคิดทั่วไปเกี่ยวกับการวัดแบบอิงเกณฑ์ แบ่งเป็น

1. แนวคิดทั่วไปเกี่ยวกับการวัดแบบอิงเกณฑ์
2. แนวคิดทั่วไปเกี่ยวกับแบบสอบอิงเกณฑ์
3. การกำหนดมาตรฐานหรือคะแนนจุดตัด
4. เกณฑ์ในการประเมินวิธีกำหนดคะแนนจุดตัด

ตอนที่ 2 แนวคิดเกี่ยวกับทฤษฎีการอ้างอิงสรุป แบ่งเป็น

1. ลักษณะของความแปรปรวน
2. คำศัพท์ที่เกี่ยวข้องกับทฤษฎีการสรุปอ้างอิง
3. ขั้นตอนของทฤษฎีการสรุปอ้างอิง
4. การศึกษาขั้นการสรุปอ้างอิง (G-Study)
5. การศึกษาขั้นการตัดสินใจ (D-Study)
6. สรุปกระบวนการวิเคราะห์ตามทฤษฎีการสรุปอ้างอิง

ตอนที่ 3 งานวิจัยที่เกี่ยวข้อง

ตอนที่ 1 แนวคิดทั่วไปเกี่ยวกับการวัดแบบอิงเกณฑ์

แนวคิดทั่วไปเกี่ยวกับการวัดแบบอิงเกณฑ์

การวัดแบบอิงเกณฑ์ (Criterion - Referenced measurement) เป็นการวัดที่ขึ้นอยู่กับเกณฑ์มาตรฐานสัมบูรณ์ที่สามารถจะบอกระดับคุณภาพของสิ่งที่จะวัดได้ และสิ่งที่ใช้เป็นมาตรฐานได้แก่ ระดับเกณฑ์ (Criterion level) คะแนนจุดตัด (Cut-off score) การผ่าน-ตก (Pass-fail) หรือ

มาตรฐานการผ่าน (Passing standard) (Glaser, 1963 อ้างถึงใน สำเริง บุญเรืองรัตน์, 2527 ; Roid and Haladyna, 1982) หรือกล่าวอีกนัยหนึ่งว่าการวัดแบบอิงเกณฑ์เป็นการวัดความสามารถของนักเรียนแต่ละคนเปรียบเทียบกับเกณฑ์ที่กำหนดไว้ (เยาวดี วิบูลย์ศรี, 2539; โกวิท ประวาลพฤกษ์, 2523) และการวัดผลแบบอิงเกณฑ์จะต้องสอดคล้องกับวัตถุประสงค์เชิงพฤติกรรมของการเรียนการสอน (สำเริง บุญเรืองรัตน์, 2527) หรือกล่าวโดยสรุปได้ว่า การวัดผลแบบอิงเกณฑ์มุ่งจะวัดความสามารถของผู้เรียนตามระดับการเรียนรู้ใน 2 ระดับดังนี้

1. วัดระดับความสามารถในการรอบรู้ (Mastery Level) คือมุ่งจะวัดความสามารถในการเรียนรู้ตามเกณฑ์มาตรฐานหรือคะแนนจุดตัดที่กำหนดไว้

2. วัดระดับความสามารถในการพัฒนา (Development Level) คือมุ่งจะวัดความสามารถในการเรียนรู้ของผู้เรียนที่สูงกว่าขั้นรอบรู้ การวัดในระดับนี้จะบ่งชี้ถึงระดับความก้าวหน้าที่ผู้เรียนได้พัฒนาขึ้นมาใหม่ จากการเรียนการสอนที่กำหนดไว้

ทอนไดรและคณะ (Thomdrike, et al., 1991) กล่าวว่าการวัดและประเมินผลแบบอิงเกณฑ์มีลักษณะที่สำคัญ คือ

1. เป็นลักษณะงาน (Task) ที่แสดงถึงความสัมพันธ์กับโดเมนที่เกี่ยวกับการสอนโดยเฉพาะ
2. มีรูปแบบที่เป็นตัวแทนและตอบสนองวัตถุประสงค์
3. มีระดับการปฏิบัติที่ยอมรับได้ว่าเป็นผู้รอบรู้ของนักเรียนแต่ละคน

จากลักษณะเฉพาะของการวัดแบบอิงเกณฑ์ดังกล่าว สามารถเปรียบเทียบลักษณะที่แตกต่างระหว่างการทดสอบแบบอิงกลุ่มกับการทดสอบแบบอิงเกณฑ์ ได้ดังต่อไปนี้

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 1 การเปรียบเทียบลักษณะการทดสอบแบบอิงกลุ่มกับการทดสอบแบบอิงเกณฑ์

ประเด็น	การทดสอบแบบอิงกลุ่ม	การทดสอบแบบอิงเกณฑ์
1.การนำไปใช้	จัดตำแหน่งการเรียนรู้ของผู้เรียนโดยเทียบกับปกติวิสัยของกลุ่ม	วัดระดับความสามารถของผู้เรียนโดยเทียบกับเกณฑ์มาตรฐานของการเรียนรู้
2. ความสำคัญ	วัดความแตกต่างของผลสัมฤทธิ์ของนักเรียนแต่ละคน	อธิบายการปฏิบัติงานเฉพาะของนักเรียน
3. การแปลความหมาย	เปรียบเทียบการปฏิบัติของนักเรียนกับคนอื่นๆ	เปรียบเทียบการปฏิบัติของนักเรียนกับโดเมนผลสัมฤทธิ์ที่เฉพาะ
4.ความครอบคลุมเนื้อหา	ครอบคลุมเนื้อหาโดยทั่วๆไปของผลสัมฤทธิ์	ครอบคลุมเนื้อหาที่เป็นลักษณะเฉพาะของการเรียนรู้
5. การวางแผนแบบสอบ	ใช้ตารางกำหนดแผนผังการสร้างข้อสอบโดยทั่วไป	ใช้ตารางกำหนดแผนผังการสร้างข้อสอบที่มีรายละเอียดเฉพาะ
6. วิธีการเลือกข้อสอบ	เลือกข้อสอบที่มีอำนาจจำแนกสูง ข้อที่ง่ายจะถูกตัดออก	ประกอบด้วยข้อสอบที่จำเป็นและเพียงพอสำหรับอธิบายการปฏิบัติ
7. มาตรฐานการปฏิบัติ	ระดับการปฏิบัติตัดสินโดยสัดส่วนที่สัมพันธ์กับกลุ่ม	ระดับการปฏิบัติตัดสินโดยมาตรฐานสัมบูรณ์

ที่มา Gronlund, N. E. (1993) How to make achievement tests and assessments. (5th ed.) Boston : Allyn and Bacon.

แบบสอบอิงเกณฑ์ (Criterion-Referenced Test) เป็นแบบสอบที่มุ่งวัดระดับการเรียนรู้ของผู้เรียนว่ามีความรู้ความสามารถอะไรบ้าง ข้อสอบอิงเกณฑ์จึงถูกสร้างให้ครอบคลุมความรู้หรือทักษะสำคัญของการเรียนรู้ที่ต้องการให้เกิดขึ้น ดังนั้นข้อมูลที่ได้จากแบบสอบอิงเกณฑ์จึงอธิบายเกี่ยวกับความรู้เฉพาะและทักษะของผู้สอบในลักษณะของคะแนน ซึ่งสามารถตีความหมายได้ว่า เป็นลักษณะงาน (Task) ที่ผู้สอบสามารถกระทำได้เมื่อเปรียบเทียบกับเกณฑ์หรือมาตรฐานที่กำหนดไว้ (บุญธรรม กิจปรีดาบริสุทธิ์, 2535; ศิริชัย กาญจนวาสี, 2535; เยาวดี วิบูลย์ศรี, 2539; Wiersma and Jurs, 1990) ผลสรุปที่ได้จากแบบสอบอิงเกณฑ์คือ การตัดสินว่ารอบรู้เมื่อนักเรียนปฏิบัติได้ถึงระดับที่ยอมรับได้หรือเกณฑ์ หรือไม่รอบรู้เมื่อนักเรียนปฏิบัติได้ไม่ถึงระดับที่ยอมรับได้หรือเกณฑ์ และแบบสอบอิงเกณฑ์จะตัดสินการปฏิบัติของนักเรียนหลังการเรียนการสอนตามวัตถุประสงค์ ซึ่งเป็นการวัด

การปฏิบัติที่เฉพาะ บางครั้งอาจเรียกการสอบแบบสอบอิงเกณฑ์ได้ว่า การสอบแบบสอบการรอบรู้ (Bott, 1996) โดย Glaser (1963 อ้างถึงใน เยาวดี วิบูลย์ศรี, 2539) กล่าวว่า แบบสอบอิงเกณฑ์จะต้องมีเกณฑ์หรือคะแนนจุดตัด (Cut-off score) เพื่อวินิจฉัยว่านักเรียนคนใดเรียนรู้ครบถ้วนตามวัตถุประสงค์หรือไม่ และเกณฑ์หรือคะแนนจุดตัดที่ใช้โดยทั่วไปควรมีคุณสมบัติ 2 ประการ คือ

1. มีลักษณะที่สามารถบรรยายถึงวัตถุประสงค์ของการเรียนการสอนได้อย่างชัดเจน

2. มีลักษณะที่สามารถบ่งชี้ ถึงประสิทธิภาพของพฤติกรรมอันเป็นที่ยอมรับได้ หรือสามารถบอกถึงระดับของการปฏิบัติที่เป็นมาตรฐาน ในขณะเดียวกันผลจากการสอบแบบอิงเกณฑ์จะต้องสามารถนำไปสู่การตีความลักษณะงานเฉพาะที่บุคคลกระทำได้อย่างอิสระ โดยไม่จำเป็นต้องอ้างอิงถึงกลุ่มปกติวิสัย (Norm) ดังจะเห็นได้ว่าคะแนนที่ได้จากแบบสอบอิงเกณฑ์มักจะ ตีความหมายเป็นเชิงสัมบูรณ์ (Absolute) มากกว่าที่จะตีความหมายในลักษณะของมาตรฐานเชิงสัมพัทธ์ (Relative standard) ดังที่ปฏิบัติกันในแบบสอบอิงกลุ่ม

ข้อควรระวังเกี่ยวกับการวัดแบบอิงเกณฑ์

เมเรน และเลแมน (Mehrens and Lehmann, 1984) ได้เสนอสิ่งที่ควรระวังสำหรับการวัดแบบอิงเกณฑ์ ไว้ดังนี้

1. คำนิยามที่ค่อนข้างเฉพาะของการวัดแบบอิงเกณฑ์คือ การประมาณสัดส่วนการปฏิบัติของนักเรียนที่สามารถบรรลุตามวัตถุประสงค์ในขอบเขตเนื้อหานั้นๆ และมีแบบสอบน้อยมากที่เป็นไปตามคำนิยามนี้ จึงควรระวังเกี่ยวกับการอ้างอิงความถูกต้องของสัดส่วนจากตัวอย่างข้อสอบไปยังความถูกต้องของสัดส่วนจากประชากรข้อสอบทั้งหมด

2. แบบสอบอิงเกณฑ์อาจจะมีความตรงตามเนื้อหาไม่มากไปกว่าแบบสอบอิงกลุ่ม

3. ถ้าเลือกข้อสอบจากพื้นฐานการตัดสินใจเกี่ยวกับคุณภาพ และความไวในการเรียนการสอนจะเป็นผลให้แบบสอบมีความลำเอียงเช่นเดียวกับการเลือกข้อสอบจากพื้นฐานด้านความสามารถระหว่างนักเรียนที่เก่งกับนักเรียนที่อ่อน

4. ในการทดสอบวัดความรอบรู้เราไม่สามารถวัดเพื่อจำแนกนักเรียนในกลุ่มเก่งได้ เพราะทั้งการสอนและการสอบอาจจะเป็นเพียงการวัดระดับความรอบรู้ขั้นต่ำ

5. แบบทสอบอิงเกณฑ์จำนวนมากจะสั้นกว่าและมีค่าความเที่ยงน้อยกว่าแบบสอบอิงกลุ่ม

6.ข้อคำถามที่ยากไปหรือการกำหนดคะแนนจุดตัดที่สูงไปอาจทำให้เกิดความคลาดเคลื่อน (Hableton and Novick, 1973 ; Millman, 1974 ; Swaminathan et. al., 1975)

7. ผลสรุปที่ได้จากแบบสอบอิงเกณฑ์ไม่ได้บอกให้ทราบว่า นักเรียนควรทำอะไรและเมื่อใดเพื่อให้บรรลุผลสัมฤทธิ์นั้นๆ (Grosswald, 1973)

8. เงื่อนไขของแบบสอบอิงเกณฑ์อาจทำให้ครูไม่คำนึงถึงการวิเคราะห์ข้อสอบเหมือนกับแบบสอบอิงกลุ่ม (Grosswald, 1973)

9. ต้องระวังในการแปลความหมายของแบบสอบอิงเกณฑ์ เพราะถ้านักเรียนตกหรือไม่ผ่านวัตถุประสงค์อาจจะมาจากความผิดพลาดของการจัดกิจกรรมการเรียนการสอน ตัวข้อสอบ คะแนนจุดตัด และ/หรือวัตถุประสงค์เองก็เป็นได้ (Ground, 1973)

แนวคิดเกี่ยวกับวิธีกำหนดมาตรฐานหรือคะแนนจุดตัด

เกณฑ์มาตรฐานหรือคะแนนจุดตัด

แนวคิดของการพัฒนาเกณฑ์มาตรฐานหรือคะแนนจุดตัด เป็นแนวคิดที่พัฒนาควบคู่มากับการพัฒนาแบบสอบอิงเกณฑ์ ทั้งนี้เนื่องจากการแปลความหมายของคะแนนจากการสอบแบบอิงเกณฑ์จะต้องกำหนดมาตรฐานขึ้นมาให้ได้ว่า จากการสอบนั้นๆผู้สอบควรจะสามารถทำข้อสอบได้ถูกต้องในระดับใด จึงถือได้ว่าผู้สอบมีสมรรถนะในเรื่องนั้นเป็นที่น่าพอใจหรือยอมรับได้ว่าเป็นผู้ที่มีความสามารถเหมาะสมในเรื่องที่มารับการทดสอบแต่ละครั้ง และถ้าผู้สอบคนใดทำแบบสอบได้คะแนนในระดับต่ำกว่าเกณฑ์ที่กำหนดก็จะถือได้ว่าเขายังไม่มีความสามารถในเรื่องนั้นๆ Glaser (1963) เรียกสภาวะของผู้สอบที่มีระดับความสามารถสูงถึงขั้นมาตรฐานที่กำหนดไว้ว่า เป็นภาวะของการ "รอบรู้" (Master Level) และเรียกภาวะตรงกันข้ามว่า "ไม่รอบรู้" (Non-Master Level)

สมหวัง พิธิยานุวัฒน์ (2541) ให้ความหมายของมาตรฐาน (Standards) ไว้ว่าหมายถึงสภาพสิ่งที่ควรจะเป็น เป็นสภาพอันประกอบด้วยคุณภาพ เช่น โรงเรียนมาตรฐาน ก็จะต้องประกอบด้วยครูที่มีความรู้ความสามารถสูง มีวัสดุอุปกรณ์ครบถ้วน บรรยากาศเอื้ออำนวยในการเรียนรู้ เป็นต้น เมื่อกล่าวถึงมาตรฐานของสิ่งใดก็หมายถึงสภาพที่เป็นอุดมการณ์อุดมคติที่ควรจะเป็นของสิ่งนั้น และเกณฑ์ (Criteria) หมายถึง สิ่งที่เราใช้ตัดสินคุณภาพของผลลัพธ์ที่ได้หรือส่วนประกอบการ (Performance) เช่นนักเรียนจะสอบผ่านถือว่ามีความรู้ความสามารถในวิชาคณิตศาสตร์ชั้น ป.3 อย่างเพียงพอจะต้องสอบได้คะแนนไม่ต่ำกว่าร้อยละ 50 เป็นต้น จะเห็นว่าเกณฑ์ที่ดีจะต้องสามารถทำนายพฤติกรรมต่อมาได้อย่างถูกต้อง เกณฑ์จะต้องมีลักษณะความเป็นมาตรฐานถ้ากำหนดเกณฑ์สูงมากกล่าวคือมีความเป็นมาตรฐานมาก ก็อาจเกิดปัญหาไม่มีทรัพยากรที่จะทำให้ทุกคนบรรลุเกณฑ์ชั้นสูงเหล่านั้น แต่ถ้าตั้งเกณฑ์ต่ำเกินไปก็จะทำให้เกิดปัญหาเรื่องคุณภาพของผู้ผ่านเกณฑ์ เพราะเกณฑ์ที่ต่ำเกินไปไม่มีอำนาจในการทำนายพฤติกรรม ระดับของเกณฑ์ที่เหมาะสมควรผันแปรระหว่างปกติวิสัยและมาตรฐาน

เยาวดี วิบูลย์ศรี (2539) เกณฑ์มาตรฐานหรือคะแนนจุดตัดตามคำนิยามเชิงปฏิบัติการ หมายถึง จุดหรือระดับค่าของคะแนนระดับใดระดับหนึ่งซึ่งอยู่ระหว่างจำนวนเต็มกับคะแนนศูนย์ ที่ผู้วัดผลใช้เป็นจุดตัดเพื่อแบ่งระดับความสามารถของผู้สอบออกเป็น "ผู้รอบรู้" กับ "ผู้ไม่รอบรู้"

แนวทางในการกำหนดมาตรฐานหรือคะแนนจุดตัด

Berk (1986 อ้างถึงใน เยาวดี, วิบูลย์ศรี,2539) ได้จัดประเภทแนวคิดของการกำหนดมาตรฐานหรือคะแนนจุดตัดไว้ 3 เกณฑ์ ดังนี้

เกณฑ์ที่ 1 แนวคิดของการกำหนดมาตรฐาน โดยอิงธรรมชาติของการเรียนรู้ในวิชาต่างๆ ซึ่งสามารถแบ่งออกเป็น 2 สภาวะ คือ

1. สภาวะตามธรรมชาติการเรียนรู้ที่สามารถแยกออกจากกันอย่างชัดเจน โดยแบ่งเป็น 2 ลักษณะคือ การเรียนรู้ได้ทั้งหมดกับการไม่ได้เรียนรู้เลยธรรมชาติเช่นนี้มีความหมายว่า ในการเรียนรู้บางวิชานั้นถ้าผู้เรียนสามารถเรียนรู้ก็จะสามารถเข้าใจได้ทั้งหมดในคราวเดียวกัน แต่ถ้ายังไม่เรียนรู้ก็จะไม่สามารถเข้าใจอะไรเลยในวิชานั้น ภาวะดังกล่าวจะปรากฏออกมาในรูปของคะแนน คือทำคะแนนจากแบบสอบเนื้อหานั้นได้ทั้งหมด หรือทำคะแนนจากแบบสอบในเนื้อหานั้นๆไม่ได้เลย ซึ่งจะแยกผู้สอบเป็นผู้รอบรู้หรือไม่รอบรู้ออกจากกันตามคุณลักษณะที่ปรากฏเป็นเกณฑ์มาตรฐาน

2. ภาวะตามธรรมชาติของการเรียนรู้อย่างต่อเนื่อง ในลักษณะที่ค่อยๆเป็นค่อยๆไป ลักษณะธรรมชาติเช่นนี้จะปรากฏออกมาในรูปของคะแนน คือสัดส่วนของคะแนนที่ทำได้ของผู้สอบจะสะท้อนให้เห็นถึงระดับการสะสมความสามารถในขอบเขตเนื้อหานั้นๆ ดังนั้นในการตัดสินว่าเป็นผู้รอบรู้หรือไม่รอบรู้ ก็จะพิจารณาได้จากระดับความสามารถช่วงใดช่วงหนึ่งของความ ต่อเนื่องในการเรียนรู้เป็นเกณฑ์มาตรฐาน

เกณฑ์ที่ 2 แนวคิดของการกำหนดมาตรฐาน โดยพิจารณาถึงองค์ประกอบพื้นฐานที่เป็นหลักในการให้ได้มาซึ่งเกณฑ์มาตรฐานนั้นๆ องค์ประกอบพื้นฐานตามเกณฑ์ข้อนี้มี 2 ส่วนหลัก คือ

1. ส่วนที่เป็นการตัดสินของผู้เชี่ยวชาญ (Judgmental) หมายความว่า มาตรการที่จะนำมากำหนดเป็นเกณฑ์มาตรฐานนั้น จะยึดเอาผลการพิจารณาจากข้อสรุปของผู้เชี่ยวชาญเฉพาะเนื้อหาในแต่ละสาขาวิชาเป็นหลัก

2. ส่วนที่เป็นข้อมูลเชิงประจักษ์ (Empirical) มาตรการที่จะนำมากำหนดเป็นเกณฑ์นั้นจะยึดเอาข้อมูลเชิงประจักษ์ที่ได้จากผลของการนำแบบสอบไปทดลองใช้ก่อนเป็นหลัก

จากองค์ประกอบพื้นฐานทั้งสองส่วน ก่อให้เกิดรูปแบบของการผสมระหว่างองค์ประกอบทั้งสอง ทำให้ได้แนวทางใหม่ในการกำหนดมาตรฐานหรือคะแนนจุดตัดอีก 3 แนวทาง คือ

ก. แนวทางที่ใช้ข้อสรุปของผู้เชี่ยวชาญเพียงอย่างเดียว

ข. แนวทางที่ใช้ข้อมูลเชิงประจักษ์เป็นส่วนสำคัญ และใช้การตัดสินของผู้เชี่ยวชาญเป็นส่วนประกอบ

ค. แนวทางที่ใช้การตัดสินของผู้เชี่ยวชาญเป็นส่วนสำคัญ และใช้ข้อมูลเชิงประจักษ์เป็นส่วนประกอบ

เกณฑ์ที่ 3 แนวคิดของการกำหนดมาตรฐานโดยการพิจารณาว่าเป็นแนวคิดใดระหว่าง 2 แนวทางต่อไปนี้

1. แนวคิดของการกำหนดค่าของเกณฑ์มาตรฐาน (Setting Standards)
2. แนวคิดของการพยายามปรับเปลี่ยนเพื่อหาเกณฑ์มาตรฐานที่เหมาะสม

วิธีกำหนดมาตรฐานหรือคะแนนจุดตัด

Hambleton (1980) และ Suen (1990) ได้รวบรวมวิธีกำหนดมาตรฐานโดยแบ่งเป็น 3 กลุ่ม ดังนี้

1. วิธีการกำหนดมาตรฐานหรือจุดตัดโดยใช้ดุลยพินิจของผู้เชี่ยวชาญ (Judgmental Methods) โดยใช้ผู้เชี่ยวชาญในเนื้อหา หรือผู้เกี่ยวข้องในเนื้อหาร่วมกันกำหนดจุดตัดขึ้นมา การกำหนดจุดตัดวิธีนี้จะได้จุดตัดที่คงที่ไม่แปรเปลี่ยนไปตามคะแนนของผู้สอบ ซึ่งเป็นการกำหนดจุดตัดแบบสัมบูรณ์ (Absolute) แบ่งเป็น 2 ลักษณะ คือ

1.1 ใช้การตัดสินข้อสอบ ได้แก่ วิธีของนีเดลสกี (Nedelsky) วิธีของแองกอฟ (Angoff) และวิธีที่พัฒนาจากวิธีแองกอฟ (Modified Angoff)

1.2 ใช้การเดา (Guessing) ได้แก่ วิธีของมิลแมน (Millman)

2. การกำหนดจุดตัดโดยอาศัยข้อมูลจากการสอบของกลุ่มผู้สอบ (Empirical Model) ผลการกำหนดจุดตัดด้วยวิธีนี้มีค่าแปรเปลี่ยน หรือสัมพันธ์กับคะแนนจากการสอบของกลุ่มผู้สอบ แบ่งเป็น

2.1 ใช้การกำหนดเกณฑ์ในการวัด (Data Criterion Measurement) ได้แก่ วิธีของลิฟวิงสตัน (Livingston) วิธีของแวน เดอร์ลินเดน และเมลเลนเบอร์ก (Van Der Linden and Mellenberge)

2.2 ใช้ทฤษฎีการตัดสินใจ (Decision Theoretic) ได้แก่ วิธีของไครวอล (Kriewal) วิธีของกลาส (Glass, 1978)

3. การกำหนดมาตรฐานจากการใช้ดุลยพินิจของผู้เชี่ยวชาญร่วมกับการใช้ข้อมูลจากการสอบของกลุ่มผู้สอบ (Combination Model) แบ่งเป็น

3.1 ใช้การตัดสินเชิงประจักษ์ (Judgmental Empirical) ได้แก่

1. ใช้การเปรียบเทียบ (Contrasting groups)

2. ใช้ระดับคาบเส้น (Borderline groups) ได้แก่ วิธีของลิฟวิงสตันและซายกี (Livingston & Zieky)

3. ใช้เกณฑ์อื่นๆประกอบ (Criterion groups) ได้แก่ วิธีหาจุดสมดุลย์ระหว่างเกณฑ์สัมบูรณ์กับเกณฑ์สัมพันธ์ของบูค (Beuk, 1986) วิธีการตัดสินโดยอาศัยสารสนเทศประกอบ(Informed Judgement)ของเบอร์ก (Berk)

3.2 พิจารณาถึงผลการศึกษาที่ตามมา (Education Consequence) ได้แก่ วิธีของ บลอก (Block)

3.3 วิธีการของเบส์ (Bayesian's Method) ได้แก่ วิธีของแฮมเบิลตันและโนวิค (Hambleton and Novic) วิธีของสคูน กูลเลี่ยน และเฟอร์รารา(Schoon, Gullion and Ferrara)

กลาส (Glass, 1978) แบ่งวิธีกำหนดมาตรฐานหรือคะแนนจุดตัดไว้ดังนี้

1. วิธีที่ใช้ผลที่ได้จากการปฏิบัติของคนอื่นมาเป็นเกณฑ์ (Performance of other Criterion) โดยกำหนดจุดตัดจากการอ้างอิงพารามิเตอร์ของประชากรผู้เข้าสอบ ซึ่งอาจกำหนดด้วยคะแนนมัธยฐานของคะแนนจากแบบสอบ หรือกำหนดตำแหน่งเปอร์เซนต์ไทล์ที่เหมาะสมขึ้นมาก่อนแล้วนำมาเทียบกับคะแนนผลสอบที่สอดคล้องกับตำแหน่งเปอร์เซนต์ไทล์นั้น แต่วิธีนี้นักทฤษฎีเกี่ยวกับแบบสอบอิงเกณฑ์เห็นว่าไม่เหมาะสมกับการวัดแบบอิงเกณฑ์

2. วิธีการนับถอยหลังจาก 100% (Counting backward from 100%) วิธีนี้จะต้องกำหนดความสำคัญของวัตถุประสงค์เชิงพฤติกรรม ถ้าจุดประสงค์ใดมีความสำคัญค่าของจุดตัดจะมีค่าสูง ถ้าจุดประสงค์ใดมีความสำคัญน้อยค่าจุดตัดก็จะลดต่ำลงมา วิธีการนับถอยหลังจาก 100% คือ นำข้อสอบและวัตถุประสงค์เชิงพฤติกรรมไปให้ผู้เชี่ยวชาญด้านนั้นๆ พิจารณาความสำคัญของวัตถุประสงค์โดยกำหนดความสำคัญเป็น 5 ระดับ คือ สำคัญมากที่สุด สำคัญมาก สำคัญปานกลาง สำคัญน้อย และสำคัญน้อยมาก โดยในแต่ละระดับคิดเป็น 95%, 90%, 85%, 80%, และ 75% ตามลำดับ หลังจากนั้นนำมาหาค่าเฉลี่ย และเปอร์เซนต์เฉลี่ยที่ได้คือ คะแนนจุดตัด

3. วิธีการเพิ่มหรือปรับคะแนนตามเกณฑ์อื่นๆ (Boot Strapping another Criterion Score) วิธีนี้อาศัยเกณฑ์ภายนอกมากำหนดว่า "สำเร็จ" หรือ "รอบรู้" โดยหาคะแนนจุดตัดจากการกระจายของคะแนนที่ได้จากการสอบของผู้ที่ถูกแบ่งเป็นผู้รอบรู้กับไม่รอบรู้ แล้วกำหนดคะแนนจุดตัดที่สามารถแบ่งผู้ที่รอบรู้ออกจากผู้ที่ไม่รอบรู้ให้สอดคล้องกับเกณฑ์ภายนอกมากที่สุด วิธีการกำหนดจุดตัดตามแนวคิดนี้ ได้แก่ Berk

4. วิธีการใช้ดุลยพินิจตัดสินความสามารถต่ำสุด (Judging minimal competency) เป็นวิธีการที่ให้ผู้เชี่ยวชาญในเนื้อหาวิชาหรือครูประจำวิชากำหนดคะแนนจุดตัด โดยการพิจารณาข้อสอบแล้วระบุว่านักเรียนที่มีความสามารถต่ำสุดที่ยอมรับได้มีคะแนนสอบผ่านเท่าไร การกำหนดคะแนนจุดตัดด้วยวิธีนี้ ได้แก่ วิธีของนีดลสกี (Nedelsky) วิธีของแองกอฟ (Angoff) และวิธีของอีเบล (Ebel) แต่วิธีนี้จะมีปัญหาเกี่ยวกับความสามารถของผู้เชี่ยวชาญในการกำหนดความสามารถชั้นต่ำสุดของผู้เรียน และความสอดคล้องหรือความเที่ยงในการตัดสินของผู้เชี่ยวชาญ

5. วิธีวิจัยเชิงปฏิบัติ (Operations research method) เป็นวิธีกำหนดคะแนนจุดตัดจากผลการทดลองที่ตามมาโดยใช้คุณลักษณะที่เกิดจากการสอนอย่างใดอย่างหนึ่ง เช่นการถ่ายโยงการเรียนรู้ ความสำเร็จในการทำงานเป็นเกณฑ์ภายนอก อาศัยโค้งการแจกแจงทางคณิตศาสตร์หรือกราฟช่วยในการตัดสินความสัมพันธ์ของคะแนนจุดตัดจากแบบสอบกับเกณฑ์ภายนอกที่กำหนดขึ้น วิธีกำหนดจุดตัดวิธีนี้ ได้แก่ วิธีของบลอก (Block)

6. วิธีใช้ทฤษฎีการตัดสินใจ (Decision theoretic procedure) วิธีนี้จะนำคณิตศาสตร์เข้ามาช่วยในการคำนวณ และใช้ในการตัดสินใจอย่างจริงจัง เช่น วิธีของเบส์ (Baysian Decision Theoretic) แต่ Gruijter, Leyden and Hambleton (1984 อ้างถึงใน อัจฉริยา ปราบอริพ่าย, 2531) เห็นว่าวิธีกำหนดจุดตัดโดยอาศัยฟังก์ชั่นความสูญเสียที่คาดหวังน้อยที่สุด ตามวิธีของเบส์ (Baysian Decision Theoretic) ไม่เหมาะที่จะนำมาใช้ในการกำหนดจุดตัด เพราะว่าการประมาณค่าคะแนนโดเมนและอัตราส่วนความสูญเสียตลอดจนการใช้โมเดลทางคณิตศาสตร์นั้นมีข้อตกลงเบื้องต้น ซึ่งในทางปฏิบัติแล้วทำได้ยากและคะแนนจุดตัดที่ได้ยังขึ้นอยู่กับกลุ่มตัวอย่างผู้สอบด้วย นอกจากนี้ยังมีวิธีการกำหนดคะแนนจุดตัดด้วยทฤษฎีการตัดสินใจของกลาส (Glass, 1978) ซึ่งเป็นวิธีที่อาศัยการข้อมูลจากการสอบกับเกณฑ์ภายนอกแบ่งผู้สอบออกเป็นกลุ่มผู้รอบรู้กับไม่รอบรู้ และพิจารณาคะแนนจุดตัดโดยอาศัยฟังก์ชันความสัมพันธ์ของคะแนนจุดตัดกับเกณฑ์ภายนอก และคะแนนจุดตัดที่ได้จะให้ค่าความคลาดเคลื่อนในการจำแนกกลุ่มผู้สอบน้อยที่สุด วิธีกำหนดคะแนนจุดตัดหรือมาตรฐานโดยวิธีใช้ทฤษฎีการตัดสินใจตามวิธีของกลาสเป็นวิธีพัฒนาจากวิธีของนีดลสกี กล่าวคือ ผู้สอบจะถูกแบ่งเป็น

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

2 กลุ่ม คือ กลุ่มผู้มีความรอบรู้และกลุ่มผู้ไม่มีความรอบรู้ โดยใช้เกณฑ์ภายนอกแทนด้วยสัดส่วน PE กับ 1-PE เมื่อทำการทดสอบผู้สอบกลุ่มนี้ด้วยแบบสอบอิงเกณฑ์และมีการกำหนดค่าหนึ่งเป็นมาตรฐาน ในที่นี้แทนด้วย CX นำข้อมูลมาทำการแจกแจงความถี่แบบสองทางจะแยกผู้สอบได้เป็น 4 กลุ่มดังภาพ (พัสวี จริตธรรม, 2538 : 38)

	เกณฑ์ภายนอก ผ่าน	ไม่ผ่าน	
แบบสอบอิงเกณฑ์ ไม่ผ่าน	PA	PB	1-Pc
ผ่าน	PC	PD	Pc
	PE	1-PE	

ตามแผนภาพ PA = สัดส่วนของผู้สอบที่ได้คะแนนต่ำกว่ามาตรฐานแต่สอบผ่านเกณฑ์ภายนอก จัดเป็นความผิดพลาดคลาดเคลื่อนในทางลบ (False negatives)

PD = สัดส่วนของผู้สอบที่ได้คะแนนสูงกว่ามาตรฐานแต่ไม่ผ่านเกณฑ์ภายนอก จัดเป็นความผิดพลาดคลาดเคลื่อนในทางลบ (False positives)

PC = สัดส่วนของผู้สอบที่ได้คะแนนสูงกว่ามาตรฐานและผ่านเกณฑ์ภายนอก

PB = สัดส่วนของผู้สอบที่ได้คะแนนต่ำกว่ามาตรฐานและไม่ผ่านเกณฑ์ภายนอก

ในกระบวนการกำหนดมาตรฐานนักวัดผลการศึกษาต้องสร้างและวิเคราะห์ข้อมูลโดยใช้ตารางแจกแจงความถี่แบบ 2 ทาง สำหรับค่าคะแนนจุดตัดหรือมาตรฐานทุกค่าที่เป็นไปได้ แล้วพิจารณาเลือกคะแนนจุดตัดหรือมาตรฐานที่เหมาะสมเป็นคะแนนที่ลดความน่าจะเป็นในการตัดสินใจผิด ซึ่งทำให้เกิดความคลาดเคลื่อน (PA+PD) ให้น้อยที่สุด และเพิ่มความน่าจะเป็นในการตัดสินใจที่ถูกต้อง (PB+PC) ให้มากที่สุด ดังสมการ

$$F(CX) = (PA+PD) / (PB+PC)$$

การกำหนดมาตรฐานวิธีนี้ จะต้องคำนวณหาค่าความน่าจะเป็นของการเกิดความคลาดเคลื่อนและการตัดสินใจที่ถูกต้อง และเลือกคะแนนมาตรฐานที่มีความน่าจะเป็นที่จะเกิดความคลาดเคลื่อนต่ำที่สุด

ตัวอย่าง โกวิทย์ ประวาลพฤกษ์ และสมศักดิ์ สินธุระเวชญ์ (2537 อ้างถึงใน พัสวี จริตธรรม, 2538) กำหนดมาตรฐานโดยใช้ทฤษฎีการตัดสินใจ จากการนำค่าระดับผลการเรียน 2.20

ขึ้นไปของรายวิชา ส 101 เป็นเกณฑ์ภายนอก และเกณฑ์ภายในของแบบสอบอิงเกณฑ์วิชาสังคมศึกษา จำนวน 10 ข้อ ในที่นี้กำหนดมาตรฐานเป็น 5, 6 และ 7 คะแนน เพื่อหามาตรฐานที่เหมาะสมที่สุดจากจำนวนนักเรียนเข้าสอบ 20 คน

เมื่อใช้มาตรฐานหรือคะแนนจุดตัด 5 คะแนน มีข้อมูลในตารางดังนี้

	เกณฑ์ภายนอก ผ่าน	ไม่ผ่าน	
ไม่ผ่าน	PA=4/20	PB=2/20	0.30
ข้อสอบอิงเกณฑ์ (5) ผ่าน	PC=6/20	PD=8/20	0.70
	0.50	0.50	1.00

F(CX) = (PA+PD) / (PB+PC)

PA = 4/20 = 0.2

PB = 2/20 = 0.1

PC = 6/20 = 0.3

PD = 8/20 = 0.4

F(CX) = (0.2+0.4) / (0.1+0.3) = 0.6/0.4 = 1.5

เมื่อใช้มาตรฐานหรือคะแนนจุดตัด 6 คะแนน มีข้อมูลในตารางดังนี้

	เกณฑ์ภายนอก ผ่าน	ไม่ผ่าน	
ไม่ผ่าน	PA=5/20	PB=2/20	0.35
ข้อสอบอิงเกณฑ์ (5) ผ่าน	PC=5/20	PD=8/20	0.65
	0.50	0.50	1.00

F(CX) = (PA+PD) / (PB+PC)

F(CX) = (5/20 + 8/20) / (2/20 + 5/20)

= (0.2+0.4) / (0.1+0.3)

= 0.6/0.4

= 1.86

เมื่อใช้มาตรฐานหรือคะแนนจุดตัด 7 คะแนน มีข้อมูลในตารางดังนี้

	เกณฑ์ภายนอก ผ่าน	ไม่ผ่าน	
ไม่ผ่าน	PA=4/20	PB=3/20	.45
ข้อสอบอิงเกณฑ์ (5) ผ่าน	PC=4/20	PD=7/20	.55
	0.50	0.50	1.00

$$F(CX) = (PA+PD) / (PB+PC)$$

$$F(CX) = (6/20 + 7/20) / (4/20 + 4/20)$$

$$= (0.30+0.35) / (0.20+0.20)$$

$$= 0.65/0.4$$

$$= 1.63$$

จากการคำนวณค่า F(CX) เมื่อคะแนนจุดตัดเป็น 5, 6, 7 คะแนน พบว่า คะแนนจุดตัด 5 คะแนนให้ค่า F(Cx) ต่ำสุด ดังนั้นคะแนนจุดตัดที่เหมาะสมที่สุดคือ 5 คะแนน

หมายเหตุ ในการที่จะกำหนดคะแนนจุดตัดตั้งแต่คะแนนต่ำสุดไปเรื่อยๆจนถึงคะแนนเต็ม แล้วพิจารณาคะแนนที่เหมาะสมที่สุด คือ คะแนนที่มีค่า F(CX) ต่ำที่สุด

เวียสมา และเจิส (Wiersma and Jurs, 1990) ได้เสนอวิธีกำหนดคะแนนจุดตัดดังนี้

1. การตัดสินของผู้เชี่ยวชาญ (Professional Judgement) วิธีนี้เป็นวิธีที่ใช้กันมากในโรงเรียน โดยผู้เชี่ยวชาญก็คือครูผู้สอนที่มีความคุ้นเคยกับเนื้อหา และระดับความสามารถของผู้เรียนในการผ่านวัตถุประสงค์ ตัวอย่างเช่น ครูอาจกำหนดว่าร้อยละ 70 ของการตอบถูกถือว่าเป็นเกณฑ์ที่เพียงพอในการสนับสนุนว่าเด็กผ่านและสามารถเริ่มเรียนในวัตถุประสงค์ใหม่ได้ วิธีนี้มีข้อจำกัดในแง่ของความเป็นอัตนัย และครูอาจมีข้อจำกัดเกี่ยวกับตัวเลือกในการกำหนดคะแนนจุดตัด เช่น อาจกำหนดเป็น 80 เปอร์เซนต์ในทุกๆวัตถุประสงค์เป็นต้น

2. วิธีของนีเดลสกี (Nedelsky Method) เป็นวิธีการกำหนดคะแนนจุดตัดที่พัฒนาจากแบบสอบทั้งหมดมากกว่าจะพิจารณาเฉพาะวัตถุประสงค์เดียว แต่ก็สามารถประยุกต์ได้กับวัตถุประสงค์ วิธีของนีเดลสกีเหมาะที่จะใช้กับแบบสอบแบบเลือกตอบในการการกำหนดระดับความสามารถขั้นต่ำสุดที่ยอมรับได้ โดยให้ผู้เชี่ยวชาญพิจารณาข้อสอบทีละข้อแล้วตัดสินว่ามีตัวเลือกใดบ้างที่นักเรียนที่มีระดับความสามารถต่ำสุดที่ยอมรับได้บอกได้ว่าผิด และคำนวณระดับการผ่านต่ำสุดได้จากผลรวมของเศษส่วนกลับของจำนวนตัวเลือกที่เหลืออยู่ เช่น ข้อสอบมี 4 ตัวเลือก ผู้เชี่ยวชาญตัดสินว่ามี 1 ตัวเลือกที่นักเรียนระดับต่ำสุดบอกได้ว่าผิด ส่วนกลับของตัวเลือกที่เหลือคือ 1/3 นำค่าเศษส่วนจากทุกข้อมารวมกันแล้วหาค่าเฉลี่ยของผลการตัดสินข้อสอบของผู้เชี่ยวชาญแต่ละคน เป็นค่าเฉลี่ยของคะแนนดิบของนักเรียนระดับต่ำสุดเพื่อนำไปหาค่าคะแนนจุดตัดต่อไป

3. วิธีของแองกอฟ (Angoff's Method) แองกอฟ (Angoff, 1971) เสนอวิธีกำหนดคะแนนจุดตัดโดยมีขั้นตอนคือ ให้ผู้เชี่ยวชาญพิจารณาข้อสอบทีละข้อแล้วประมาณค่าความน่าจะเป็นที่กลุ่มนักเรียนที่มีความสามารถขั้นต่ำสุดในระดับที่ยอมรับได้ จะตอบข้อสอบข้อนั้นถูกและคะแนนจุดตัดจะมีค่าเท่ากับค่าเฉลี่ยของการประมาณค่าข้อสอบของผู้เชี่ยวชาญทั้งหมด

4. วิธีการเปรียบเทียบระหว่างกลุ่มของผู้สอบ (Contrasting Group) วิธีนี้เสนอโดยนีเดลสกี (Nedelsky,1954) เป็นวิธีที่ขึ้นอยู่กับการปฏิบัติการสอบของกลุ่ม โดยการให้กลุ่มผู้เชี่ยวชาญที่คุ้นเคยกับนักเรียนที่จะทำแบบสอบ แบ่งผู้สอบออกเป็น 2 กลุ่มคือ กลุ่มที่รอบรู้ และกลุ่มที่ไม่รอบรู้ แล้วให้นักเรียนทั้งสองกลุ่มทำแบบสอบ นำผลสอบของทั้งสองกลุ่มมาเขียนกราฟแจกแจงความถี่ของคะแนนร่วมกัน คะแนนจุดตัดที่ได้จะเท่ากับคะแนนตรงจุดตัดของกราฟระหว่างกลุ่มที่รอบรู้กับกลุ่มที่ไม่รอบรู้

อีเบล และฟิสบาย (Ebel and Frisbie, 1986) เสนอวิธีกำหนดคะแนนจุดตัดไว้ดังนี้

1. การกำหนดคะแนนจุดตัดต่ำกว่า 100 เปอร์เซนต์ของการตอบถูก เช่น อาจกำหนดเป็น 75, 70 หรือ 65 เปอร์เซ็นต์ วิธีนี้มีข้อจำกัดคือเป็นการกำหนดร้อยละของการผ่านตามอำเภอใจ และข้อสอบอาจจะมีความยากและง่ายไป เมื่อนักเรียนทำแบบสอบผ่านหรือตกอาจจะเนื่องมาจากลักษณะคำถามของข้อสอบมากว่ามาจากระดับความสามารถที่แท้จริงของนักเรียน

2. วิธีของอีเบล (Ebel's Method) วิธีนี้เสนอให้พิจารณาถึงลักษณะของข้อสอบในแง่ของความเกี่ยวข้อง (relevance) และค่าความยาก (difficulty) ของข้อสอบแต่ละข้อในแบบสอบ โดย

ความเกี่ยวข้องแบ่งเป็น 4 ระดับคือ จำเป็น (Essential) สำคัญ(Important) ยอมรับได้ (Acceptable) และไม่แน่ใจ (Questionable) ส่วนค่าความยากมี 3 ระดับ คือ ง่าย ปานกลาง และยาก แล้วให้ผู้ตัดสินเสนอเปอร์เซนต์ที่คาดว่าผู้สอบที่มีความสามารถต่ำสุดที่ยอมรับได้น่าจะสอบผ่าน (คล้ายกับโอกาสในการตอบถูก)

3. วิธีนี้มีข้อตกลงว่า ความสามารถของผู้สอบเป็นสิ่งเกี่ยวข้องที่จำเป็น และการทำแบบสอบจะเป็นตัวเลือกผู้ที่มีความสามารถสูง และปฏิเสธผู้ที่มีความสามารถต่ำ นั่นคือแบบสอบยังเป็นพื้นฐานที่สำคัญแต่คะแนนผ่านไม่เท่ากับสัดส่วนของการตอบถูก แต่เท่ากับคะแนนสูงกว่า 50 เปอร์เซนต์ 60 เปอร์เซนต์ หรือ 90 เปอร์เซนต์ ของคะแนนที่ตกอยู่ นั่นคือ ค่าเฉลี่ยต่ำสุดประมาณ 1 ใน 2 หรือ 1 ใน 3 และ 1 ใน 10 ของผู้สอบต้องตก

4. วิธีที่ 4 เป็นการรวมวิธีที่ 2 และวิธีที่ 3 มีหลักในการตัดสินคะแนนผ่าน โดยการพิจารณาเปอร์เซนต์ของการตอบถูกและเปอร์เซนต์ของผู้สอบควบคู่กันไป

5. เป็นวิธีที่ใช้การปฏิบัติการสอบของผู้สอบเป็นพื้นฐานในการตัดสินคะแนนผ่าน

เกณฑ์ในการประเมินวิธีกำหนดคะแนนจุดตัด

Berk (1986) เสนอเกณฑ์การประเมินวิธีกำหนดคะแนนจุดตัดไว้ดังนี้

1. ความพอเพียงตามหลักวิชาการ (Technical adecacy) เกณฑ์เกี่ยวกับความพอเพียง ตามหลักวิชาการได้มาจาก 3 แหล่งข้อมูล คือ

ก. คุณลักษณะของวิธีกำหนดมาตรฐานที่ตัดสินโดยผู้เชี่ยวชาญที่มีหลักฐานทางด้านการวิจัยและการวิเคราะห์เชิงเหตุผล

ข. เป็นมาตรฐานที่ตรงกับปัญหา และได้รับคำแนะนำจากมาตรฐานการทดสอบด้านการศึกษาและจิตวิทยา (Standards for Educational and Psychological Testing)

ค. การตัดสินใจที่ว่าด้วยประเด็นเกี่ยวกับการกำหนดมาตรฐาน (Standard Setting) จาก 3 แหล่งข้อมูลนี้นำมากำหนดเป็นเกณฑ์การประเมินวิธีกำหนดจุดตัดได้ 6 เกณฑ์ ดังนี้

1. เป็นวิธีให้สารสนเทศเกี่ยวกับการแบ่งกลุ่มที่เหมาะสม โดยวิธีที่ใช้ควรให้จุดตัดที่แบ่งกลุ่มผู้สอบออกเป็นผู้รอบรู้กับไม่รอบรู้ ผ่านกับไม่ผ่าน หรือมีความสามารถกับไม่มีความสามารถ

2. ควรเป็นวิธีที่คำนึงถึงการปฏิบัติของข้อสอบและคุณภาพของข้อสอบ โดยเป็นวิธีที่มีความไวต่อระดับความยากของข้อสอบที่ต่างกัน ตลอดจนคุณลักษณะอื่นๆที่เกี่ยวกับข้อสอบ เช่นอำนาจจำแนกของข้อสอบระหว่างกลุ่มที่รอบรู้กับไม่รอบรู้

3. เป็นวิธีที่ไวต่อการเรียนการสอน สอดคล้องกับการเรียนการสอน เพราะถ้าแบบสอบวัดได้ไม่ตรงตามวัตถุประสงค์การเรียนการสอน คะแนนจุดตัดที่ได้ก็จะไม่มีความตรงและไม่เหมาะสม

4. เป็นวิธีที่ใช้สถิติอย่างเหมาะสมและถูกต้องตามวัตถุประสงค์ ตลอดจนการแปลความหมายได้ถูกต้อง

5. เป็นวิธีที่สามารถระบุมาตรฐานที่แท้จริงให้เห็นความแตกต่างระหว่างมาตรฐานบนสเกลของคะแนนจริง (มาตรฐานที่แท้จริง) กับสเกลของคะแนนที่สังเกตได้

6. เป็นวิธีที่มีหลักฐานเกี่ยวกับความตรงในการตัดสินจุดตัด

2. เกณฑ์เกี่ยวกับการนำไปปฏิบัติได้จริง (Practicability) เกณฑ์ความเพียงพอด้านวิชาการทั้ง 6 ข้อยังไม่เพียงพอสำหรับการเลือกวิธีกำหนดมาตรฐาน จึงเพิ่มเกณฑ์เกี่ยวกับการนำไปปฏิบัติได้จริง ซึ่งว่าด้วยเรื่อง การนำไปใช้ การคำนวณ และการตีความหมาย ดังนี้

1. เป็นวิธีที่ง่ายต่อการนำไปปฏิบัติ มีความชัดเจน เข้าใจได้ง่าย มีขั้นตอนในการปฏิบัติที่ไม่ยุ่งยากซับซ้อน ไม่เสียเวลา และเป็นระบบ

2. เป็นวิธีที่ใช้สถิติในการคำนวณที่ง่าย หรือสามารถใช้โปรแกรมสำเร็จรูปทางสถิติได้ทั้งไมโครคอมพิวเตอร์ (Microcomputer) และเมนเฟรม (Mainframe)

3. เป็นวิธีที่แปลความหมายได้ง่าย และสามารถอธิบายให้ผู้ที่เกี่ยวข้อง เช่น นักเรียน ครู ผู้ปกครอง หรือผู้บริหารโรงเรียน เข้าใจได้

4. เป็นวิธีที่น่าเชื่อถือ บางครั้งวิธีที่ใช้สถิติที่ยุ่งยากก็ไม่ใช่วิธีน่าเชื่อถือเสมอไป เพราะอาจทำให้เกิดความสับสนได้มากกว่าวิธีที่ใช้การคำนวณอย่างง่ายๆ

เบอร์ก (Berk, 1986) ใช้เกณฑ์การประเมินวิธีกำหนดจุดตัดจำนวน 10 ข้อเกี่ยวกับความเพียงพอตามหลักวิธีดำเนินการ และการนำมาปฏิบัติดังกล่าวข้างต้น มาประเมินวิธีกำหนดคะแนนจุดตัดทั้ง 23 วิธี ยกเว้นเฉพาะวิธีที่ใช้ข้อมูลในการทดลองสอบประมาณความคลาดเคลื่อนในการจำแนกผู้สอบออกเป็นผู้รอบรู้ไม่จริง และผู้ไม่รอบรู้ไม่จริง ซึ่งเป็นวิธีที่ยุ่งยากซับซ้อนและนำมาปฏิบัติได้ยาก โดยสรุปผลการประเมินดังนี้

ตารางที่ 2 ผลการประเมินวิธีกำหนดคะแนนจุดตัด 23 วิธี

วิธีกำหนดคะแนนจุดตัด	เกณฑ์การประเมิน										ค่าเฉลี่ยการประเมิน
	วิธีดำเนินการ						การนำมาปฏิบัติ				
	1	2	3	4	5	6	7	8	9	10	
ก. วิธีที่ใช้ดุลยพินิจของผู้เชี่ยวชาญตัดสิน											
1.adjusted/modified M-C Angoff	++	- -	-	++	++	- -	++	-	++	++	+
2. Angoff	++	- -	-	++	++	- -	+	++	++	++	+
3. Angoff - Nedelsky combination	++	- -	-	-	+	- -	- -	-	-	+	-
4. Difficulty - importance estimate	++	- -	-	-	- -	- -	-	+	+	+	-
5. Difficulty - Relavance Ebel	++	- -	-	-	++	- -	-	+	-	-	-
6. Difficulty - Ebel	++	+	-	-	-	- -	-	+	-	-	-
7. Item specification	++	- -	-	-	- -	- -	+	+	+	+	-
8. Modified M-C Angoff	++	- -	-	-	-	- -	++	++	++	++	+
9. Nedelsky	++	- -	-	-	++	- -	- -	-	-	+	-
10. Relavance-taxomony Ebel	++	- -	- -	-	-	- -	-	+	-	-	-
11. Two-choice	++	- -	-	-	-	- -	++	+	+	+	- +
ข. วิธีที่ใช้ดุลยพินิจของผู้เชี่ยวชาญร่วมกับข้อมูลจากการสอบ (Judgmental-empirical)											
1. Absolute - relative compromise I	++	+	+	++	+	- -	-	- -	-	-	+ -
2. Absolute - relative compromise II	++	+	+	++	+	- -	-	- -	-	-	+ -

ตารางที่ 2 (ต่อ) ผลการประเมินวิธีกำหนดคะแนนจุดตัด 23 วิธี

วิธีกำหนดคะแนนจุดตัด	เกณฑ์การประเมิน										ค่าเฉลี่ยการประเมิน
	วิธีดำเนินการ						การนำมาปฏิบัติ				
	1	2	3	4	5	6	7	8	9	10	
3. Angoff - contrasting group plus composite	++	++	+	+	+	++	- -	- -	+	+	+
4. Informed judgment	++	+	+	++	++	- -	+	++	++	++	+/++
5. Iterative Angoff	++	+	+	+	++	- -	- -	++	+	++	+
6. Iterative two - choice Angoff	++	+	+	+	+	- -	- -	-	+	++	+
7. Modified Angoff - empirical	++	+	-	-	- -	- -	+	+	++	++	+
ค. วิธีที่อาศัยข้อมูลจากการสอบในการตัดสิน (Empirical-judgmental)											
1. Borderline group	++	++	+	+	-	- -	+	+	+	-	+
2. Contrasting groups	++	++	++	+	-	++	-	+	+	-	+
3. Criterion groups	++	++	++	+	-	++	-	+	+	-	+
4. Educational consequences	++	++	++	- -	- -	- -	+	+	-	-	-+
5. Norm-referenced criterion	++	++	-	-	- -	-	+	+	-	-	-+

หมายเหตุ - - หมายถึง ไม่เหมาะสมอย่างยิ่ง

- หมายถึง ไม่เหมาะสม

\+ หมายถึง เหมาะสม

\+ + หมายถึง เหมาะสมอย่างยิ่ง

ที่มา Berk, R. A. (1986) A Consumer's Guide to Setting Performance Standards on Criterion-Referenced Tests. Review of Educational Research, 56, 137 - 172.

จากการประเมินวิธีกำหนดจุดตัดทั้ง 23 วิธีของ Berk พบว่า

1. วิธีกำหนดจุดตัดที่ใช้ดุลยพินิจของผู้เชี่ยวชาญ เป็นวิธีที่ไม่มีหลักการที่ชัดเจนขึ้นอยู่กับคุณสมบัติของผู้เชี่ยวชาญ แต่พบว่าวิธีของแองกอฟเป็นวิธีที่เหมาะสมมีความสมดุลย์ทั้งวิธีการและการนำไปปฏิบัติ

2. วิธีกำหนดจุดตัดที่ใช้ดุลยพินิจของผู้เชี่ยวชาญร่วมกับข้อมูลจากการสอบในการตัดสิน พบว่า วิธีใช้ดุลยพินิจตัดสินโดยอาศัยสารสนเทศประกอบ เป็นวิธีที่เหมาะสมที่สุดในการกำหนดจุดตัดทั้ง 23 วิธี

3. สำหรับวิธีกำหนดจุดตัดที่อาศัยข้อมูลจากการสอบในการตัดสิน พบว่า วิธีการเปรียบเทียบระหว่างกลุ่มผู้สอบเป็นวิธีที่เหมาะสมตามเกณฑ์

เกณฑ์สำหรับการกำหนดจุดตัดด้วยวิธีใช้ดุลยพินิจของผู้เชี่ยวชาญ

เบอร์ก (Berk, 1986) เสนอเกณฑ์สำหรับการกำหนดคะแนนจุดตัด โดยใช้ดุลยพินิจของผู้เชี่ยวชาญ ดังนี้

1. การเลือกผู้เชี่ยวชาญ ตลอดจนประเด็นของการสุ่มตัวอย่าง
2. การฝึกอบรมผู้เชี่ยวชาญ
3. กระบวนการในการตัดสินข้อสอบ แบบสอบ หรือผู้สอบ
4. รูปแบบในการตัดสินใจของผู้เชี่ยวชาญ (กระบวนการกลุ่มหรือเป็นอิสระ)
5. วิธีการในการรวมข้อมูลเข้าด้วยกัน
6. วิธีการวิเคราะห์การตัดสินใจ และการตัดสินใจร่วมกับข้อมูลจากการปฏิบัติ

องค์ประกอบในการกำหนดคะแนนจุดตัด

มิลแมน (Millman, 1973 อ้างถึงใน Hambleton, et al., 1978) เสนอองค์ประกอบในการกำหนดคะแนนจุดตัด ดังนี้

1. คะแนนของผู้สอบคนอื่น (performance of others) การกำหนดจุดตัดที่พิจารณาองค์ประกอบด้านนี้ ทำโดยกำหนดจุดตัดที่พิจารณาจากจำนวนนักเรียนที่จะสอบผ่านไว้ก่อน หรือใช้ระดับคะแนนของผู้ที่ได้รับการรับรองแล้วว่ามีความรู้ในเนื้อหานั้นๆมากำหนดจุดตัด

2. เนื้อหาของข้อสอบ (Item Content) โดยให้ผู้เชี่ยวชาญพิจารณาตัดสินว่าข้อสอบแต่ละข้อในแบบสอบนั้นผู้สอบตอบได้อย่างถูกต้องหรือไม่

3. ผลทางการศึกษาที่ตามมา (Educational Consequences) เป็นการพิจารณาว่า คะแนนจุดตัดเท่าใดจึงจะทำให้เกิดประโยชน์ทางการศึกษามากที่สุด นั่นคือถ้าความรู้หรือทักษะที่วัดเป็นพื้นฐานของสิ่งที่จะเรียนต่อไปควรตั้งเกณฑ์ให้สูง เพราะถ้ากำหนดคะแนนจุดตัดต่ำเกินไปนักเรียนที่ผ่านก็อาจไม่สามารถเรียนบทเรียนต่อไปได้อย่างมีประสิทธิภาพ แต่ถ้ากำหนดคะแนนจุดตัดสูงเกินไปก็จะทำให้นักเรียนไม่ผ่านต้องเสียเวลาในการเรียนซ่อมเสริมโดยไม่จำเป็น

4. คุณค่าทางจิตวิทยาและทางเศรษฐกิจ (Psychological and Financial) คุณค่าทางจิตวิทยาและทางเศรษฐกิจในที่นี้รวมทั้ง แรงจูงใจ ความเบื่อ การทำลายความรู้สึกส่วนตัว เงินและเวลา การกำหนดคะแนนจุดตัดควรพิจารณาว่า ถ้าคุณค่าเหล่านี้ไม่มากนักจุดตัดควรสูง หรือถ้าเกิดผลทางลบสำหรับนักเรียนที่สอบผ่าน เช่น เกิดความสับสนหรือประสิทธิภาพในการเรียนลดลงก็ควรกำหนดจุดตัดให้สูงเช่นกัน

5. ความคลาดเคลื่อนจากการเดาและการสุ่มข้อสอบ (Error Due to Guessing and Item Sampling) การประมาณความสามารถของนักเรียนจะมีความคลาดเคลื่อนที่เป็นระบบ (Systematic error) เมื่อรูปแบบของข้อสอบในแบบสอบเปิดโอกาสให้นักเรียนตอบถูกโดยการเดา การกำหนดคะแนนจุดตัดให้สูงจะช่วยลดการเดาที่คาดว่ามีอยู่ในการทำแบบสอบ

ข้อเสนอแนะเกี่ยวกับคะแนนจุดตัด

แฮมเบิลตัน (Hambleton, 1978) ให้ข้อเสนอแนะเกี่ยวกับการกำหนดคะแนนจุดตัด ดังต่อไปนี้

1. การพิจารณาคะแนนจุดตัดควรทำโดยกลุ่มคนที่ทำงานร่วมกันหลายๆกลุ่ม เช่น ครู ผู้บริหาร ผู้ปกครอง ผู้เชี่ยวชาญด้านหลักสูตรและนักเรียน สำหรับจำนวนคนในแต่ละกลุ่มขึ้นกับความสำคัญของแบบสอบที่นำมาพิจารณา และจำนวนของตารางวิเคราะห์โดเมนของข้อสอบ

2. ผู้ตัดสินควรนำตารางวิเคราะห์โดเมนของข้อสอบมาอภิปรายร่วมกันด้วย

3. การใช้วิธีของอีเบลหรือนีเดลสกี ควรทดลองกับกลุ่มตัวอย่างหลายๆกลุ่มเพื่อตัดสินและแก้ปัญหาเกี่ยวกับความแตกต่างระหว่างกลุ่ม

4. ตารางวิเคราะห์โดเมนข้อสอบที่มีความซับซ้อน หรือมีความสำคัญมาก ควรให้เวลาในการพิจารณามากขึ้น และควรกำหนดจุดตัดให้สูงกว่า

5. พยายามทำให้เกิดความมั่นใจว่าผู้ตัดสินมีความรู้เกี่ยวกับการใช้แบบสอบ และรู้ว่าจะใช้กับนักเรียนกลุ่มใด

6. เมื่อใช้ผู้ตัดสินตั้งแต่ 2 กลุ่มหรือมากกว่า ควรศึกษาความคงที่ในการกำหนดน้ำหนักและความแตกต่างในการกำหนดน้ำหนักของผู้ตัดสินและอภิปรายเพื่อหาการตัดสินที่สอดคล้องกัน

7. นำข้อมูลเกี่ยวกับการสอบครั้งก่อนมาพิจารณาเพื่อกำหนดจุดตัด

8. เปรียบเทียบสภาพการรอบรู้ของผู้สอบที่ยังไม่ได้เรียนและเรียนมาแล้ว หากมีนักเรียนจำนวนมากได้รับการจัดประเภทผิดต้องพิจารณาคะแนนจุดตัดอีกครั้ง

9. ถ้าหลักสูตรมีการเปลี่ยนแปลง ต้องมีการตรวจสอบคะแนนจุดตัดใหม่ให้เหมาะสม

ความตรงเกี่ยวกับการตัดสินการรอบรู้ (Validity of mastery decision)

Suen (1990) กล่าวว่าในการใช้คะแนนจุดตัดเพื่อแบ่งผู้เรียนเป็นกลุ่มผู้รอบรู้และไม่รอบรู้ จะต้องพิจารณาถึงความตรงหรือความถูกต้องในการตัดสินแบ่งกลุ่มด้วย เพื่อดูว่าผู้ที่จัดอยู่ในกลุ่มรอบรู้นั้นมีประสิทธิภาพจริงหรือไม่ ซึ่งเป็นคำถามเกี่ยวกับความตรงตามเกณฑ์สัมพันธ์นั่นเอง (Criterion-relate validity) และถ้าวิธีกำหนดคะแนนจุดตัดต่างๆที่ใช้ตัดสินกลุ่มผู้รอบรู้และไม่รอบรู้สามารถพิสูจน์ได้ เราก็สามารถนำมาใช้เป็นเกณฑ์ในการวัดได้ และความตรงตามเกณฑ์สัมพันธ์ก็คือ ความสัมพันธ์ระหว่างการตัดสินการรอบรู้จากการทำแบบสอบกับการตัดสินการรอบรู้บนพื้นฐานของเกณฑ์การวัด เช่นคะแนนการสอบใช้ทำนายประสิทธิภาพของการปฏิบัติและ ประสิทธิภาพของการปฏิบัติที่ได้จากการตัดสินของผู้เชี่ยวชาญเป็นเกณฑ์การวัด หรือใช้คะแนนการสอบของผู้เรียนที่มีวัตถุประสงค์การศึกษาที่เฉพาะเป็นเกณฑ์การวัด และการแบ่งกลุ่มผู้สอบออกเป็น 4 กลุ่มตามวิธีการใช้ทฤษฎีการตัดสินใจ (Decision Theoretic Approaches) ถ้าค่าในกลุ่มผู้รอบรู้และไม่รอบรู้ทั้งสองเกณฑ์มีค่าสูงแสดงถึงความตรงตามเกณฑ์สัมพันธ์ในการตัดสินใจ

ตอนที่ 2 แนวคิดเกี่ยวกับทฤษฎีการสรุปอ้างอิง (Generalizability Theory)

ครอนบาค และคณะ (Cronbach,Gleser,Nanda and Rajaratnam,1972) ได้เสนอหลักการเกี่ยวกับทฤษฎีการสรุปอ้างอิง (Generalizability Theory) ขึ้นเป็นครั้งแรกและพัฒนาเพื่อประมาณค่าความเที่ยงของแบบสอบในสถานการณ์ของการสอบต่างๆและสามารถอธิบายแหล่งความคลาดเคลื่อนต่างๆได้มากขึ้น ในขณะที่ทฤษฎีการวัดแบบดั้งเดิมอธิบายความคลาดเคลื่อนเป็นค่าคงที่ค่าเดียวไม่สามารถแบ่งแยกได้ นอกจากนี้ทฤษฎีการสรุปอ้างอิงไม่ต้องมีข้อตกลงเบื้องต้นเกี่ยวกับแบบสอบคู่ขนาน (Nunnally, 1978) และสามารถประมาณค่าความคลาดเคลื่อนได้ครอบคลุมเงื่อนไขการวัดโดยใช้การวิเคราะห์ความแปรปรวน (Analysis of Variance) ทฤษฎีการสรุปอ้างอิงทำให้ได้ค่าสัมประสิทธิ์ความเที่ยงหรือสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจสัมบูรณ์ และค่าสัมประสิทธิ์ความเที่ยงหรือสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์

ลักษณะของความแปรปรวนของคะแนน

หลักการทางสถิติที่เกี่ยวข้องกับทฤษฎีการสรุปอ้างอิง คือ ความแปรปรวนและค่าเฉลี่ยของคะแนนทั้งหมด แต่เนื่องจากการใช้ค่าเฉลี่ยเมื่อแปลความหมายระหว่างคะแนนของผู้สอบเกิดความลำเอียงในการเป็นตัวแทน เพราะในสถานการณ์จริงไม่มีใครได้คะแนนเท่ากับค่าเฉลี่ย ความแปรปรวนเป็นดัชนีที่บอกผลรวมของคะแนนความคลาดเคลื่อนจากค่าเฉลี่ย และชี้ให้เห็นความคลาดเคลื่อนในการใช้ค่าเฉลี่ยเป็นตัวแทนของคะแนนผู้สอบ นั่นคือค่าความแปรปรวนน้อยแสดงว่าค่าเฉลี่ยอธิบายคะแนนของผู้สอบได้ถูกต้อง ถ้าค่าความแปรปรวนมากแสดงว่าค่าเฉลี่ยอธิบายคะแนนของผู้สอบไม่ถูกต้อง การทราบขนาดของความแปรปรวนเป็นประโยชน์ต่อการวัดความคลาดเคลื่อนและทำให้ทราบถึงที่มาของการขาดความเที่ยงด้วย

แหล่งของความแปรปรวน

ข้อมูลที่ได้จากการสอบมีแหล่งความแปรผันที่แตกต่างกันขึ้นอยู่กับสถานการณ์ในการสอบ ดังตัวอย่างต่อไปนี้

สถานการณ์ที่ 1 ให้นักเรียนทำข้อสอบ 1 ข้อ ผู้ตรวจให้คะแนน 6 คะแนนจากคะแนนเต็ม 10 คะแนน ในสถานการณ์นี้เราไม่ทราบว่าคะแนน 6 คะแนน เป็นตัวแทนของคะแนนที่แท้จริงของนักเรียน เพราะมีข้อมูลไม่เพียงพอในการประมาณค่าความเที่ยง แต่เมื่อนำข้อสอบไปให้ผู้ตรวจให้คะแนนในวันต่อมา ปรากฏว่าผู้ตรวจให้คะแนนเท่าเดิม เราสรุปได้ว่าผู้ตรวจมีความคงที่ในการตัดสินใจนั่นคือมีความสอดคล้องในการตัดสินของผู้ตรวจ (Intrajudge consistency) ถ้าในแต่ละวันให้คะแนนไม่เท่ากันแสดงว่าไม่มีความสอดคล้องในการตัดสินของผู้ตรวจ การวัดเกี่ยวกับความสอดคล้องในการตัดสินของผู้ตรวจคือความแปรปรวนของการให้คะแนนในแต่ละครั้งของผู้ตัดสินคนเดียว ถ้าความแปรปรวนมากแสดงว่ามีความสอดคล้องในการตัดสินของผู้ตรวจต่ำ และมีความคลาดเคลื่อนในการตัดสินมาก

สถานการณ์ที่ 2 ให้ผู้ตรวจหลายคนซึ่งแต่ละคนมีความสอดคล้องในการตัดสินของผู้ตรวจ ให้คะแนนนักเรียนที่ทำข้อสอบเหมือนกัน ในการให้คะแนนข้อสอบเดียวกันในเวลาหนึ่ง เราจะได้คะแนนจากผู้ตรวจแต่ละคน และค่าความแปรปรวนที่ได้ไม่ใช่ความคลาดเคลื่อนในการตัดสินของผู้ตรวจ แต่เป็นความคลาดเคลื่อนระหว่างผู้ตรวจ ถ้าค่าความแปรปรวนน้อยแสดงว่า มีความสอดคล้องในการตัดสินระหว่างผู้ตรวจสูง

สถานการณ์ที่ 3 ให้ผู้ตรวจ 1 คนที่มีมีความสอดคล้องในการตัดสินของผู้ตรวจ (Intrajudge consistency) ให้คะแนนนักเรียน 1 คนที่ทำข้อสอบหลายข้อ ในแต่ละครั้งปรากฏว่า

คะแนนต่างกัน แสดงว่ามีความคลาดเคลื่อนของคะแนน (ข้อสอบ) ความคลาดเคลื่อนดังกล่าวไม่ใช่ความคลาดเคลื่อนในการตัดสินของผู้ตรวจ หรือความคลาดเคลื่อนระหว่างผู้ตรวจ แต่เป็นความคลาดเคลื่อนระหว่างข้อสอบชี้ให้เห็นถึงการขาดความสอดคล้องภายในข้อสอบจากผู้สอบคนเดียวกัน

สถานการณ์ที่ 4 นำแบบสอบที่มีความสอดคล้องภายในข้อสอบให้นักเรียนทำการสอบและให้ผู้ตรวจที่มีความสอดคล้องในการตัดสินของผู้ตรวจและความสอดคล้องในการตัดสินระหว่างผู้ตรวจเป็นผู้ตรวจให้คะแนนทั้งหมด ปรากฏว่านักเรียนได้คะแนนต่างกัน แสดงว่ามีความแปรปรวนของคะแนนที่เกิดระหว่างผู้สอบ

ความแปรปรวนจริงและความคลาดเคลื่อน

ความแปรปรวนระหว่างผู้ตรวจ ความแปรปรวนภายในของผู้ตรวจ และความแปรปรวนระหว่างข้อสอบ เป็นความแปรปรวนที่ทำให้ขาดความเที่ยงหรือความสอดคล้อง นั่นคือถ้ามีความแปรปรวนเหล่านี้มาก แสดงว่าความคลาดเคลื่อนมาตรฐานในการวัดก็จะสูง เรียกความแปรปรวนทั้งสามแหล่งนี้ว่า ความแปรปรวนของความคลาดเคลื่อน (Error Variance) สำหรับความแปรปรวนระหว่างผู้สอบไม่ถือว่าเป็นความคลาดเคลื่อนมาตรฐานของการวัด แต่จะเป็นตัวที่ชี้ให้เห็นถึงการกระจายของคะแนนหรือความแตกต่างของผู้สอบ โดยเรียกความแปรปรวนระหว่างผู้สอบนี้ว่า ความแปรปรวนจริง (Actual variance)

ความแปรปรวนของคะแนนตามทฤษฎีการสรุปอ้างอิง

สำหรับทฤษฎีการสรุปอ้างอิง คะแนนแอกภาพ (Universe score ; Mp) คือค่าเฉลี่ยของคะแนนการวัดซ้ำหลายๆครั้งตามเงื่อนไขการวัดในแอกภาพของการสรุปอ้างอิง ค่าความคลาดเคลื่อนของการวัด ถูกจำแนกออกเป็นความคลาดเคลื่อนจากฟาเซทหรือกลุ่มเงื่อนไขของการวัด และความคลาดเคลื่อนจากแหล่งที่เหลืออื่นๆและการวัดแต่ละครั้งไม่จำเป็นต้องใช้แบบสอบคู่ขนาน เหมือนทฤษฎีการทดสอบแบบดั้งเดิม ส่วนความแปรปรวนของคะแนนที่สังเกตได้เป็นผลรวมของความแปรปรวนของคะแนนแอกภาพกับความแปรปรวนของคะแนนความคลาดเคลื่อนจากฟาเซทหรือองค์ประกอบต่างๆของการวัด และความแปรปรวนของความคลาดเคลื่อนแหล่งอื่นๆ ดังนี้

$$X_{pi} = T_{pi} + E_{pi} + e_{pi}$$

$$\sigma^2_{X_p} = \sigma^2_{\mu_p} + \sigma^2_{E_i} + \sigma^2_{e_p}$$

ความแปรปรวนของคะแนนความคลาดเคลื่อนจากองค์ประกอบต่างๆของการวัด แบ่งออกเป็น 2 ประเภท ได้แก่

1. ความแปรปรวนของคะแนนความคลาดเคลื่อนสัมบูรณ์ (Absolute Error Variance ; σ^2_{ABS}) หรือ σ^2_Δ คือความแปรปรวนของ $\mu_p - X_p$ ซึ่งคำนวณได้จากผลรวมของความแปรปรวนของคะแนนจากแหล่งต่างๆยกเว้น $\sigma^2_{\mu_p}$ หรือ σ^2_p

2. ความแปรปรวนของคะแนนความคลาดเคลื่อนสัมพัทธ์ (Relative Error Variance ; σ^2_{REL}) หรือ σ^2_δ คือความแปรปรวนของ $\mu_p - X_p$ ที่คำนวณได้จากผลรวมของความแปรปรวนของคะแนนจากแหล่งต่างๆที่มีปฏิสัมพันธ์ร่วมกับผู้สอบ (p)

คำศัพท์ที่เกี่ยวข้องกับทฤษฎีการสรุปอ้างอิง

1. ประชากร (Population) หมายถึง สิ่งที่มุ่งวัดทั้งหมดในสถานการณ์ของการสอบ ได้แก่ บุคคลหรือผู้สอบ

2. ฟาเซทของการวัด (Facet) หมายถึงชุดหรือกลุ่มของเงื่อนไขของการวัดที่มีลักษณะเดียวกัน เป็นองค์ประกอบที่คาดว่ามีผลต่อความคลาดเคลื่อนของการวัด เช่น ความยาวของข้อสอบ วิธีการวัด จำนวนครั้งของการสอบ จำนวนผู้ตรวจให้คะแนน เป็นต้น ฟาเซทที่ต้องการศึกษาอาจจะเป็นองค์ประกอบสุ่มหรือองค์ประกอบที่เจาะจง ถ้าเงื่อนไขการวัดถูกเลือกมาอย่างเจาะจงจากองค์ประกอบที่ศึกษา ผู้ศึกษาสามารถทำการสรุปอ้างอิงไปยังองค์ประกอบเฉพาะในระดับของเงื่อนไขที่เลือกมาศึกษาเท่านั้น แต่ถ้าเงื่อนไขได้รับการสุ่มเพื่อเป็นตัวแทนองค์ประกอบที่ศึกษา ผู้ศึกษาสามารถทำการสรุปอ้างอิงไปยังระดับต่างๆขององค์ประกอบที่ศึกษาได้

3. เงื่อนไขของการวัด (Condition of measurement) เป็นระดับของฟาเซทของการวัด เช่น ฟาเซทของจำนวนผู้ตรวจอาจกำหนดจำนวนระดับเป็น 1, 2 และ 3 คน ฟาเซทของความยาวแบบสอบอาจกำหนดความยาวของแบบสอบเป็น 10, 20 และ 30 ข้อ เป็นต้น

4. เอกภพ (Universe) หมายถึง เงื่อนไขของการวัดที่สนใจทั้งหมดของแต่ละฟาเซท(Facet) เช่น จำนวนข้อสอบทั้งหมด จำนวนผู้ตรวจทั้งหมด

เอกภพของค่าที่ได้จากการสังเกตทั้งหมด (admissible observation) เป็นกลุ่มเงื่อนไขของการวัดที่เป็นไปได้ ซึ่งสามารถวัดหรือสังเกตได้จากการเก็บข้อมูลการทดสอบของประชากรภายใต้เงื่อนไขของการวัดทั้งหมดในแต่ละฟาเซท เช่น ฟาเซทของข้อสอบ ฟาเซทของผู้ตรวจ

เอกภพของการอ้างอิง (Universe of generalization) เป็นเงื่อนไขการวัดทั้งหมดที่เป็นเป้าหมายของการสรุปอ้างอิง และเป็นการวัดที่ครอบคลุมเงื่อนไขที่สนใจทั้งหมด

สัมประสิทธิ์การสรุปอ้างอิงหรือสัมประสิทธิ์ความเที่ยง

สัมประสิทธิ์การสรุปอ้างอิง (Generalizability coefficient : ρ^2_p) เป็นอัตราส่วนระหว่างความแปรปรวนของคะแนนเอกภาพ และความแปรปรวนของคะแนนที่สังเกตได้ ดังสมการต่อไปนี้

$$G \text{ - coefficient} = \frac{\sigma^2_p}{\sigma^2_p + \text{Error Variance}}$$

สัมประสิทธิ์การสรุปอ้างอิง (Generalizability coefficient : ρ^2_p) เป็นดัชนีที่ใช้อธิบายความคงที่ของการวัดเช่นเดียวกับค่าสัมประสิทธิ์ความเที่ยงแบบดั้งเดิม สามารถใช้คำนวณช่วงของความเชื่อมั่นของคะแนนเอกภาพ หรือใช้ในสมการถดถอยในการประมาณค่าคะแนนเอกภาพและใช้ในการปรับแก้ค่าสหสัมพันธ์ที่ลดลงอันเนื่องมาจากความคลาดเคลื่อน

สัมประสิทธิ์การสรุปอ้างอิง สามารถอธิบายในรูปอัตราส่วนระหว่างความแปรปรวนของคะแนนเอกภาพกับคะแนนที่สังเกตได้ (Cronbach, et al., 1972) เช่น สัมประสิทธิ์การสรุปอ้างอิงมีค่าเท่ากับ .90 แสดงว่าความแตกต่างที่วัดได้ร้อยละ 90 เป็นความแตกต่างอันเนื่องมาจาก คะแนนเอกภาพ และเป็นความแตกต่างอันเนื่องมาจากความคลาดเคลื่อนคิดเป็นร้อยละ 10

เนื่องจากความแปรปรวนของคะแนนความคลาดเคลื่อนจากองค์ประกอบต่างๆของการวัดแบ่งเป็น 2 ประเภท จึงทำให้ค่าสัมประสิทธิ์การสรุปอ้างอิงประกอบด้วย 2 ประเภท ดังนี้

1. สัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ (Generalizability coefficient for absolute decision) สัมประสิทธิ์นี้บ่งบอกความเที่ยงหรือความสอดคล้องในสถานการณ์ของการตัดสินใจที่ขึ้นกับคะแนนผู้สอบตามลำพัง ไม่มีการเปรียบเทียบภายในกลุ่มหรือระหว่างกลุ่ม เช่น แบบสอบอิงเกณฑ์ซึ่งใช้ตรวจความสามารถของผู้สอบกับเกณฑ์ที่กำหนดไว้ล่วงหน้า

2. สัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์ (Generalizability coefficient for relative decision) สัมประสิทธิ์นี้บ่งบอกความเที่ยงหรือความสอดคล้องในสถานการณ์การตัดสินใจที่มีการเปรียบเทียบคะแนนระหว่างผู้สอบ เช่นแบบสอบอิงกลุ่มซึ่งใช้เปรียบเทียบคะแนนระหว่างผู้สอบด้วยการเข้าโค้งการแจกแจงคะแนนและตัดเกรด หรือให้ระดับผลการเรียน หรือการสอบแข่งขันเพื่อหาผู้ที่ได้คะแนนสูงเข้าศึกษา

ถึงแม้ว่าค่าสัมประสิทธิ์การสรุปอ้างอิง จะมีความหมายเช่นเดียวกับความเที่ยงตามทฤษฎีการวัดแบบดั้งเดิม แต่ยังมีประเด็นที่แตกต่างกันดังนี้

1. การวัดแต่ละครั้งมีค่าสัมประสิทธิ์การสรุปอ้างอิงได้มากกว่า 1 ค่า

2. การอ้างอิงไปยังเอกภพใดจะต้องระบุและอธิบายเอกภพนั้นให้ชัดเจน และต้องสุ่มเงื่อนไขนั้นมาศึกษาด้วย

3. ค่าสัมประสิทธิ์การสรุปอ้างอิงสามารถบอกถึงความเป็นเอกพันธ์ของคะแนนเอกภพได้ ถ้าข้อสอบที่นำมาศึกษาเป็นตัวอย่างสุ่มจากเอกภพ ข้อสอบชุดที่มีความเป็นเอกพันธ์จะสามารถใช้คะแนนสังเกตแทนคะแนนเอกภพได้อย่างมั่นใจ

ขั้นตอนของทฤษฎีการสรุปอ้างอิง

ทฤษฎีการสรุปอ้างอิง ประกอบด้วยขั้นตอนที่สำคัญ 2 ขั้นตอน คือ การศึกษาเพื่อการสรุปอ้างอิง หรือการศึกษา G (Generalizability study or G - Study) และการศึกษาเพื่อการตัดสินใจ หรือการศึกษา D (Decision Study or D - Study)

การศึกษาเพื่อการสรุปอ้างอิง (G -Study) เป็นการสรุปอ้างอิงผลที่ได้จากการศึกษาตัวอย่างการวัดตามเงื่อนไขที่สนใจ บรรยายความแปรปรวนของความคลาดเคลื่อนจากแหล่งต่างๆ เพื่อสรุปอ้างอิงไปยังเอกภพของการวัด จุดประสงค์ของการศึกษาเพื่อการสรุปอ้างอิงคือ ต้องการประมาณค่าความแปรปรวนของคะแนนจริง และความแปรปรวนของคะแนนความคลาดเคลื่อนจากแหล่งความคลาดเคลื่อนต่างๆที่สนใจ และใช้เป็นข้อมูลสำหรับวางแผนเพื่อตัดสินใจ ดังนั้นการออกแบบการศึกษาเพื่อการสรุปอ้างจึงควรครอบคลุมเงื่อนไขของการวัดที่ต้องการตัดสินใจนำแบบสอบไปใช้ในการศึกษาเพื่อการตัดสินใจ

การศึกษาเพื่อการตัดสินใจ (D - Study) เป็นการใช้ข้อมูลจากการศึกษาเพื่อการสรุปอ้างอิงที่สอดคล้องกับจุดประสงค์เฉพาะของการวัด ตัดสินใจเลือกใช้แบบสอบ หรือวิธีวัดในสถานการณ์ต่างๆของการวัด

1. การศึกษาเพื่อการสรุปอ้างอิง (G -Study)

การศึกษาเพื่อการสรุปอ้างอิง เป็นกระบวนการที่มุ่งประมาณความแปรปรวนจากแหล่งต่างๆภายใต้เอกภพของคะแนนที่สังเกตได้ โดยมีขั้นตอนดังนี้

1. กำหนดสิ่งที่จะวัด เช่น ความสามารถด้านการเขียนของนักเรียน

2. กำหนดฟาเซทของการวัด เช่น กำหนดให้มี 2 ฟาเซท คือ ข้อสอบ และผู้ตรวจให้คะแนน

3. กำหนดเอกภพของการสังเกต เช่น ฟาเซทของข้อสอบ ฟาเซทของผู้ตรวจให้คะแนน

4. กำหนดความสัมพันธ์ระหว่างฟาเซท ว่าจะเป็นแบบ Crossed (แทนด้วย X) หรือ Nested (แทนด้วย :) ถ้าต้องการศึกษาแบบให้ผู้ตรวจทุกคนตรวจข้อสอบทุกข้อ จะมีรูปแบบ

Crossed (I X r) แต่ถ้าให้ผู้ตรวจแต่ละคนตรวจชุดข้อสอบต่างกันก็จะเป็นแบบ Nested (i : r) ทั้งนี้แล้วแต่ความสนใจของผู้ศึกษา ซึ่งโดยหลักการแล้วถ้าเป็นแบบ Crossed จะทำให้ได้แหล่งความแปรปรวน (Variance Component) มากกว่าแบบ Nested ซึ่งมีประโยชน์ที่จะทำให้ทราบค่าว่าความแปรปรวนแหล่งใดมีผลต่อสัมประสิทธิ์การสรุปอ้างอิง และช่วยให้สามารถเลือกรูปแบบที่เหมาะสมในขั้น D-Study ตลอดจนได้ค่าสัมประสิทธิ์การสรุปอ้างอิงมากกว่า nested ในกรณีที่มีรูปแบบเป็น 1 ฟาเซท โดยสรุปแล้วในขั้น G-Study ควรใช้แบบ Crossed เพราะขั้นนี้มุ่งประมาณค่าความแปรปรวนจากแหล่งต่างๆอย่างเต็มที่ สำหรับในขั้น D-Study นั้นขึ้นอยู่กับความสนใจในการศึกษา

5. การเก็บรวบรวมข้อมูล โดยการสุ่มตัวอย่างมาศึกษา เช่น สุ่มตัวอย่างนักเรียน n_p คน ตัวอย่างข้อสอบ n_i ข้อ และตัวอย่างผู้ตรวจ n_r คน

6. การวิเคราะห์ข้อมูล หลังจากผ่านขั้นตอนในการพิจารณารูปแบบการวัด และได้นำไปสู่การเก็บรวบรวมข้อมูลจากรูปแบบที่กำหนดแล้ว นำข้อมูลที่ได้ไปวิเคราะห์ทางสถิติเพื่อประมาณขนาดของความคลาดเคลื่อนในการวัดจากแหล่งต่างๆ กระบวนการทางสถิติที่ใช้ในทฤษฎีการสรุปอ้างอิงคือ การวิเคราะห์ความแปรปรวน (Analysis of Variance (ANOVA) Procedure) เพื่อหาแหล่งความแปรปรวนของคะแนน ตัวอย่างเช่น การศึกษา G ในรูปแบบ Cross design (p x I) จะมีแหล่งความแปรปรวนของคะแนน 3 แหล่ง คือ ผู้สอบ (p) ข้อสอบ (r) และปฏิสัมพันธ์ระหว่างผู้สอบกับข้อสอบ (pi) หรือส่วนที่เหลือ (residual) สามารถคำนวณผลรวมความเบี่ยงเบนกำลังสอง (Mean of Squares : SS) ได้ดังนี้

ผลรวมความเบี่ยงเบนกำลังสองของผู้สอบ $(SS_p) = n_i \sum(Xp. - X..)^2$

ผลรวมความเบี่ยงเบนกำลังสองของข้อสอบ $(SS_i) = n_p \sum(X.i - X..)^2$

ผลรวมความเบี่ยงเบนกำลังสองของส่วนที่เหลือ $(SS_r) = SSt - SSp - SSi$

โดยที่ $SSt = \sum_i \sum_p (Xpi - X..)^2$

เมื่อ Xp. = คะแนนเฉลี่ยของการทำข้อสอบ 1 ข้อของคนที่ p

X.i = คะแนนเฉลี่ยข้อที่ I

X.. = คะแนนเฉลี่ยของผู้เข้าสอบทุกคน (grand mean)

SSt = ผลรวมความเบี่ยงเบนกำลังสองทั้งหมด

และชั้นแห่งความเป็นอิสระ (degree of freedom : df) มีค่าเท่ากับ

$df_p = N - 1$ เมื่อ N = จำนวนผู้สอบ

$df_i = K - 1$ เมื่อ K = จำนวนข้อสอบ

$df_{pi} = (N - 1)(K - 1)$

ดังนั้นค่าเฉลี่ยกำลังสอง (Mean Squares : MS) มีค่าเท่ากับ

$$MS_p = SS_p / (N - 1)$$

$$MS_i = SS_i / (K - 1)$$

$$MS_{pi} = SS_{pi} / (N - 1)(K - 1)$$

ในการวิเคราะห์ความแปรปรวนโดยทั่วไปนั้น ประมาณค่าเฉลี่ยกำลังสอง (MS) เพื่อทดสอบความมีนัยสำคัญของสัดส่วนความแปรปรวน โดยการใช้สถิติทดสอบ F (F-Test) และใช้ F-ratio ในการตัดสินว่าจะคงไว้หรือจะปฏิเสธสมมติฐานนั้น แต่ในการประยุกต์การวิเคราะห์ความแปรปรวนกับทฤษฎีการสรุปอ้างอิงนั้น F-ratio เป็นสิ่งที่ไม่มีความจำเป็น เพราะต้องการเพียงค่าเฉลี่ยกำลังสอง (Mean Squares - MS) จากการวิเคราะห์ความแปรปรวนมาเป็นฐานในการประมาณองค์ประกอบของความแปรปรวนเท่านั้น โดยที่องค์ประกอบของความแปรปรวนสามารถประมาณค่าได้จากค่าเฉลี่ยกำลังสอง ดังนี้

$$\sigma^2_p = (MS_p - MS_{pi}) / K$$

$$\sigma^2_i = (MS_i - MS_{pi}) / N$$

$$\sigma^2_{pi} = MS_{pi}$$

เมื่อ σ^2_p = ค่าประมาณองค์ประกอบความแปรปรวนของผู้สอบ

σ^2_i = ค่าประมาณองค์ประกอบความแปรปรวนของข้อสอบ

σ^2_{pi} = ค่าประมาณองค์ประกอบความแปรปรวนของส่วนที่เหลือ

ดังนั้นการวิเคราะห์ความแปรปรวนในที่นี้เป็นเพียงกระบวนการเพื่อการประมาณค่าความแปรปรวนในขั้นการศึกษาเพื่อสรุปอ้างอิง ไม่ใช่เป็นการทดสอบสมมติฐานเหมือนที่ใช้ในทางสถิติโดยทั่วไป สำหรับความแปรปรวนที่ประมาณได้ในการศึกษาเพื่อสรุปอ้างอิง (Estimated G-Study Variance Component) เป็นการประมาณค่าความแปรปรวนที่แท้จริง (Actual Variance of Parameter) กับเงื่อนไข (ข้อสอบหรือผู้ตรวจ) ในเอกภพของการสังเกตที่ยอมรับได้และกับบุคคลเดียว (Single person) ในประชากรตัวอย่าง เช่น σ^2_p เป็นค่าประมาณซึ่งในการตีความจะพิจารณานักเรียนแต่ละคนในประชากรนักเรียน โดยนักเรียนแต่ละคนจะได้คะแนนจากการทดสอบทุกข้อ N_i คะแนนในเอกภพของการสังเกตที่ยอมรับได้ ดังนั้นนักเรียนแต่ละคนจะได้คะแนนเฉลี่ย (μ_p) และความแปรปรวนของคะแนนเฉลี่ยของนักเรียนคือ $\sigma^2(\mu_p)$ ซึ่งก็คือ σ^2_p สำหรับความแปรปรวนของผลหลักอื่นๆ ในฟาเซทข้อกระทงหรือฟาเซทของผู้ตรวจก็ตีความหมายในทำนองเดียวกัน

2. การศึกษาเพื่อการตัดสินใจ (D-Study)

การศึกษาเพื่อการตัดสินใจ (D-Study) เป็นขั้นตอนการนำผลที่ได้จากการประมาณความแปรปรวนในการศึกษาเพื่อสรุปอ้างอิง (G-Study) มาตัดสินใจเลือกวิธีการวัดที่เฉพาะเจาะจงและเหมาะสมเพื่อให้ได้ค่าประมาณความเที่ยงที่ดี ในการศึกษาการตัดสินใจ (D-Study) มีขั้นตอนดังนี้

1. กำหนดเอกภพของการสรุปอ้างอิง (Universe of generalizability) ที่ผู้ศึกษาต้องการสรุปอ้างอิง ซึ่งเอกภพของการสรุปอ้างอิงอาจประกอบด้วยเงื่อนไขทั้งหมด หรือเป็นเพียงเงื่อนไขย่อยในเอกภพของการสังเกตที่ยอมรับได้ ในขั้นนี้ผู้วิจัยต้องเลือกว่าจะใช้โมเดลใด จะเป็นโมเดลเชิงสุ่ม (Random Model) โมเดลกำหนด (fixed Model) หรือโมเดลผสม (Mix Model) ซึ่งโดยหลักการแล้ว ควรใช้โมเดลแบบสุ่มดีที่สุดเพราะจะทำให้การสรุปอ้างอิงได้กว้างขวางกว่า โมเดลแบบผสมและโมเดลแบบกำหนดตามลำดับ

2. กำหนดขนาดตัวอย่าง (Sample Size) จำนวนเงื่อนไขของตัวอย่างฟาเซทในการศึกษาเพื่อสรุปอ้างอิง (G-Study) ไม่จำเป็นต้องเป็นตัวอย่างในการศึกษาเพื่อตัดสินใจ (D-Study)ทั้งหมด ทั้งนี้การกำหนดขนาดตัวอย่างขึ้นอยู่กับความสนใจของผู้ศึกษา

3. โครงสร้างของแบบที่ศึกษา (Design Structure) ใน D-Study อาจใช้โครงสร้างของแบบที่ศึกษาเหมือนหรือต่างจากในการศึกษาเพื่อสรุปอ้างอิง (G-Study) ก็ได้ เช่น ถ้าในการศึกษาเพื่อตัดสินใจ (D-Study) ตัดสินใจให้นักเรียนทุกคนทำข้อสอบทุกข้อเหมือนกัน จะมีโครงสร้างในลักษณะ Cross design

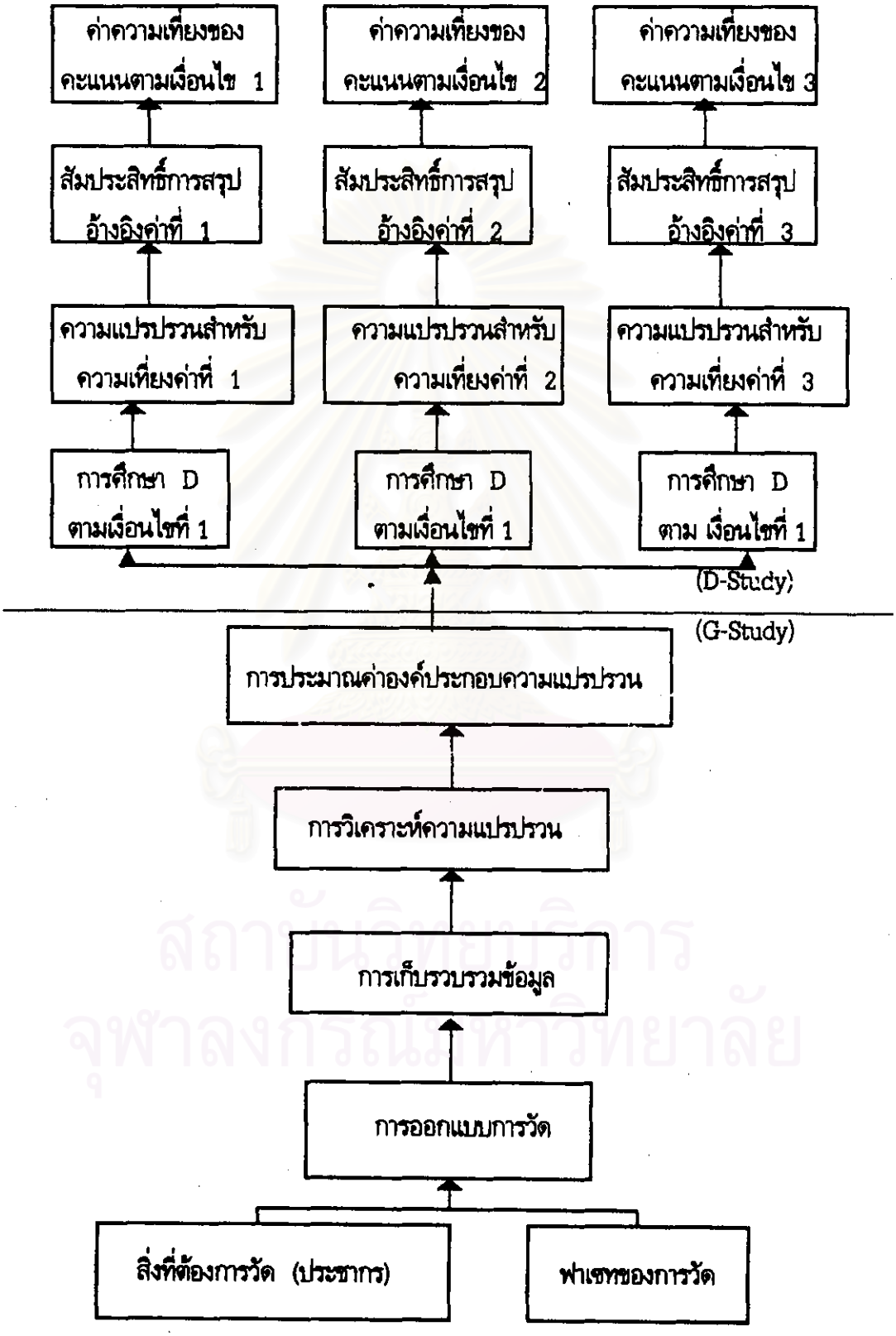
4. การประมาณค่าความแปรปรวน (Estimated D study Variance Component) มีขั้นตอนดังนี้

4.1 ระบุรูปแบบและขนาดกลุ่มตัวอย่าง

4.2 คำนวณหา universe score variance และ error score variance แล้วคำนวณค่าสัมประสิทธิ์การสรุปอ้างอิง

ทั้งนี้ได้สรุปขั้นตอนเพื่อแสดงกระบวนการการวิเคราะห์ตามทฤษฎีการสรุปอ้างอิง ดังภาพต่อไปนี้ (Suen, 1990)

ภาพที่ 1 สรุปกระบวนการของทฤษฎีการสรุปอ้างอิง



ตอนที่ 3 งานวิจัยที่เกี่ยวข้อง

จากการศึกษางานวิจัยที่เกี่ยวข้องกับการกำหนดมาตรฐานหรือคะแนนจุดตัด ผู้วิจัยแบ่งงานวิจัยออกเป็น 3 กลุ่มดังต่อไปนี้

1. การเปรียบเทียบคะแนนจุดตัดของแบบสอบอิงเกณฑ์ โดยใช้วิธีกำหนดคะแนนจุดตัดเหมือนกัน เช่น ชมภู จันทรอมรพร (2522) กำหนดคะแนนจุดตัดของแบบสอบอิงเกณฑ์วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 1 จำนวน 4 ฉบับ โดยใช้วิธีของเบส์ พบว่าคะแนนจุดตัดที่ได้มีค่าไม่เท่ากันแต่ใกล้เคียงกัน โดยแบบสอบที่มีคะแนนจุดตัดสูงจะเป็นข้อสอบที่ง่ายกว่าฉบับอื่น ซึ่งสอดคล้องกับการกำหนดเกณฑ์โดยพิจารณาความยากง่ายของแบบสอบ กล่าวคือถ้า แบบสอบง่ายควรกำหนดเกณฑ์ไว้สูงกว่าแบบสอบที่ยาก องค์ประกอบที่มีผลต่อค่าคะแนนจุดตัดโดยอาศัยกระบวนการเชิงทฤษฎีการตัดสินใจของเบส์ค่อนข้างมากคือ ค่าคะแนนจุดตัดเบื้องต้นจากวิธีการที่ใช้ในการหาค่าความสูญเสียของการตัดสินให้สอบผ่านหรือสอบตก

พูนศักดิ์ พระรัตน์ภูมี (2535) เปรียบเทียบคุณภาพและคะแนนจุดตัดของแบบสอบที่เขียนจากจุดประสงค์เชิงพฤติกรรมและฟอร์มข้อสอบ พบว่า ค่าความยาก อำนาจจำแนก และความเที่ยงของแบบสอบไม่ต่างกันอย่างมีนัยสำคัญทางสถิติ และแบบสอบทั้งสองมีคะแนนจุดตัดเท่ากันคือ 11 คะแนน

Beuk (1984) เสนอวิธีหาจุดสมดุลระหว่างเกณฑ์สัมบูรณ์กับเกณฑ์สัมพัทธ์ โดยให้ผู้เชี่ยวชาญเสนอเกณฑ์สัมบูรณ์ในรูปของเปอร์เซ็นต์ของข้อสอบน้อยที่สุดที่ผู้สอบควรตอบถูก และเกณฑ์สัมพัทธ์อยู่ในรูปของเปอร์เซ็นต์ของผู้สอบที่คาดว่าจะสอบผ่าน หลังจากนั้นนำมาหาความสัมพันธ์ระหว่างเกณฑ์สัมบูรณ์กับเกณฑ์สัมพัทธ์ และนำแบบสอบไปทดสอบกับนักเรียนเพื่อหาความสัมพันธ์ระหว่างคะแนนสอบกับเปอร์เซนต์ของผู้สอบที่จะเป็นผู้สอบได้ และนำความสัมพันธ์ทั้งสองดังกล่าวมาพล็อตกราฟ จุดตัดของสมการทั้งสองนั่นคือ จุดสมดุลระหว่างเกณฑ์สัมบูรณ์กับเกณฑ์สัมพัทธ์

2. การเปรียบเทียบคุณภาพของแบบสอบอิงเกณฑ์เมื่อใช้วิธีกำหนดคะแนนจุดตัดต่างกัน ได้แก่ รังสรรค์ มณีเล็ก (2527) เปรียบเทียบความเที่ยงของแบบสอบอิงเกณฑ์ โดยใช้วิธีการกำหนดคะแนนจุดตัดที่ต่างกัน 4 วิธี คือ วิธีนับถอยหลังจาก 100% (Counting blackward from 100%) วิธีของนีเดลสกี (Nedelsky) วิธีของเบอร์ก (Berk) และวิธีของเบส์ (Bayesian) พบว่า วิธีการกำหนดคะแนนจุดตัดที่ได้คะแนนจุดตัดสอดคล้องกันทั้ง 3 กลุ่มตัวอย่าง คือ วิธีนับถอยหลังจาก 100% จะให้ค่าคะแนนจุดตัดสูงสุด และวิธีของนีเดลสกีซึ่งมีคะแนนจุดตัดต่ำสุด สำหรับค่าความเที่ยงของแบบสอบจากการใช้วิธีการกำหนดจุดตัดที่ต่างกันนั้น ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

พัสวี จริตธรรม (2538) พัฒนาและตรวจสอบคุณภาพของเกณฑ์การประเมินการออกเสียง "ร" และ "ล" ในภาษาไทย โดยใช้การกำหนดมาตรฐาน 3 วิธี คือ วิธีกำหนด ตามผลการปฏิบัติของผู้อื่น วิธีการนับลดจาก 100% และวิธีการกำหนดมาตรฐานตามทฤษฎีการตัดสินใจ ผลการวิจัยได้มาตรฐานสำหรับการประเมินในการอ่านออกเสียง การสัมภาษณ์ และรวมการอ่านและการสัมภาษณ์ เท่ากับคะแนนร้อยละ 77, 47 และ 79 ตามลำดับ และความเที่ยงของผู้ประเมินเท่ากับ .89, .92 และ .95 ตามลำดับ

3. การเปรียบเทียบวิธีการกำหนดคะแนนจุดตัด ได้แก่ อัจฉริยา ปราบอริพ่าย (2531) เปรียบเทียบวิธีการกำหนดจุดตัดของแบบสอบ และค่าดัชนีความคงที่ของการตัดสินคะแนนจุดตัด ซึ่งตัดสินโดยผู้เชี่ยวชาญที่ไม่มีสารสนเทศและผู้เชี่ยวชาญที่มีสารสนเทศพบว่าการให้และไม่ให้สารสนเทศแก่ผู้เชี่ยวชาญมีผลทำให้คะแนนจุดตัดแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และพบว่าดัชนีความคงที่ของการตัดสินคะแนนจุดตัดที่ตัดสินโดยผู้เชี่ยวชาญที่ไม่มีสารสนเทศและผู้เชี่ยวชาญที่มีสารสนเทศ ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

กัญจนา สินทรัพนศิริกุล (2534) เปรียบเทียบคะแนนจุดตัด จำนวนผู้รอบรู้ และความเที่ยงในการจำแนกผู้สอบเป็นผู้รอบรู้ โดยการประยุกต์ทฤษฎีการตอบสนองข้อสอบตามวิธีของแวนเดอร์ เลน เดน และ เดน เมื่อใช้วิธีกำหนดเกณฑ์มาตรฐานด้วยวิธีของนีเดลสกี และวิธีกลุ่มเด็กคาบเส้น พบว่า คะแนนจุดตัดที่ได้จากวิธีนีเดลสกีและวิธีกลุ่มเด็กคาบเส้นเท่ากัน ทั้งในกลุ่มนักเรียนที่มีความสามารถระดับดีและพอใช้ เมื่อใช้วิธีของแวนเดอร์ เลน เดน และคะแนนจุดตัดที่ได้จากวิธีนีเดลสกี วิธีของเดน

ครอสและคณะ (Cross, et al.,1984) ศึกษาความสอดคล้องในการกำหนดคะแนนจุดตัด (Indices of the consistency) ระหว่างวิธีของแองกอฟ วิธีของนีเดลสกี และวิธีของเจเกอร์ โดยให้ผู้เชี่ยวชาญกำหนดคะแนนจุดตัด 3 ครั้ง และวิเคราะห์ความสอดคล้องในการกำหนดคะแนนจุดตัดด้วยการประยุกต์ตามทฤษฎีการสรุปอ้างอิง (Generalizability Theory) พบว่า วิธีของแองกอฟมีความสอดคล้องในการกำหนดคะแนนจุดตัดสูงที่สุด ส่วนวิธีของนีเดลสกีมีความสอดคล้องในการกำหนดคะแนนจุดตัดต่ำสุด

อาราสมิธ (Arrasmith, 1987 อ้างถึงใน อัจฉริยา ปราบอริพ่าย, 2531) ศึกษาเกี่ยวกับความคลาดเคลื่อนในการกำหนดคะแนนจุดตัดด้วยวิธีของแองกอฟ และวิธีการเปรียบเทียบระหว่างกลุ่มโดยอาศัยทฤษฎีการสรุปอ้างอิง ทฤษฎีการตอบสนองข้อสอบ และ bootstrap estimation พบว่าการหาค่าดัชนีความคงที่โดยทฤษฎีการตอบสนองข้อสอบ และการวิเคราะห์ความแตกต่างของความน่าจะเป็นในการตอบข้อสอบได้ถูกระหว่างการประมาณโดยใช้ดุลยพินิจของผู้เชี่ยวชาญในการกำหนด

คะแนนจุดตัดด้วยวิธีของแองกอฟ สำหรับทฤษฎีการสรุปอ้างอิง และ bootstrap estimation สามารถประมาณค่าความคลาดเคลื่อนรวมๆในการตัดสินของผู้เชี่ยวชาญได้

เบรนเนนและล็อกวูด (Brennan and Lockwood, 1980) ศึกษาเปรียบเทียบการกำหนดจุดตัดด้วย วิธีของแองกอฟและนีเดลสกี โดยใช้ทฤษฎีการสรุปอ้างอิง ดำเนินการวิจัยโดยใช้กลุ่มตัวอย่างเป็นผู้ตัดสิน 5 คน ตัดสินข้อสอบ 126 ข้อ ด้วยวิธีของแองกอฟ และวิธีของนีเดลสกี พบว่า วิธีของแองกอฟ และวิธีของนีเดลสกี ให้คะแนนจุดตัดแตกต่างกันอย่างมีนัยสำคัญทางสถิติ และความเบี่ยงเบนมาตรฐานของคะแนนจุดตัดของวิธีนีเดลสกีจะเป็น 2 เท่าของวิธีของแองกอฟ

เพลคและอิมพารา (Plake and Impara, 1996) ศึกษาความความสอดคล้องในการตัดสินคะแนนจุดตัดด้วยวิธีของแองกอฟโดยให้ผู้เชี่ยวชาญกำหนดสัดส่วนการตอบข้อสอบได้ถูกต้องของผู้สอบที่มีความสามารถต่ำสุดที่ยอมรับได้ โดยใช้ข้อสอบ 24 ข้อและให้ผู้เชี่ยวชาญตัดสิน 2 ครั้ง ผลสรุปชี้ให้เห็นว่า มีความความสอดคล้องในการตัดสินคะแนนจุดตัดของผู้เชี่ยวชาญสูง ในกรณีแสดงความแตกต่างของค่าเฉลี่ยสัมบูรณ์สำหรับการประมาณค่าการปฏิบัติข้อสอบ และแสดงให้เห็นว่าคุณสมบัติของแบบสอบมีผลต่อระดับความสอดคล้องในการตัดสินคะแนนจุดตัด

นอร์ซินีและคณะ (Norcini, et al., 1987) ศึกษาผลของช่วงเวลาในการตัดสินที่ต่างกัน คือ ก่อนการประชุม ระหว่างการประชุม และหลังการประชุม ต่อการกำหนดคะแนนจุดตัดด้วยวิธีแองกอฟ พบว่า คะแนนจุดตัดที่ได้จากการตัดสินใจของผู้เชี่ยวชาญ 3 ช่วงเวลาดังกล่าว ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

Cizek and Fitzgerald (1996) ศึกษาเปรียบเทียบการตัดสินคะแนนจุดตัดของผู้เชี่ยวชาญด้วยวิธีแองกอฟโดยใช้กระบวนการกลุ่มกับให้ผู้เชี่ยวชาญตัดสินอย่างอิสระ พบว่า คะแนนจุดตัดไม่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

อิมพาราและเพลค (Impara and Plake, 1998) ตรวจสอบความสามารถของครูในการประมาณค่าสัดส่วนของการทำข้อสอบถูกของนักเรียน 2 กลุ่มคือ กลุ่มนักเรียนที่คาบเส้น (Borderline group)และกลุ่มนักเรียนทั้งหมด พบว่าการประมาณค่าของครูไม่สอดคล้องกับความยากของข้อคำถามที่ได้จริงในกลุ่มนักเรียนที่คาบเส้นและกลุ่มนักเรียนทั้งหมด โดยคะแนนของนักเรียนจะสูงกว่าการประมาณค่าของครู และค่าความถูกต้องของการประมาณไม่แปรเปลี่ยนตามประสบการณ์ในการสอนของครู จากการศึกษาความสัมพันธ์ระหว่างค่าความยากที่แท้จริงกับค่าประมาณมีค่าเท่ากับ 0.78

Shepard (1994) พบว่า ผู้ตัดสินจะประมาณค่าความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องของผู้เรียนสูงกว่าคะแนนความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องของผู้เรียนที่แท้จริงสำหรับข้อสอบที่ยาก และประมาณค่าความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องของผู้เรียนได้ต่ำกว่าคะแนนความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องของผู้เรียนที่แท้จริงสำหรับข้อสอบที่ง่าย

อิมพาราและเพลค (Impara and Plake, 1997) ได้เปรียบเทียบวิธีการกำหนดคะแนนจุดตัดระหว่างวิธีของแองกอฟแบบดั้งเดิม (The traditional Angoff's method)กับวิธีของอิมพาราและเพลค หรือวิธีตัดสินว่าได้หรือตก (The yes-no method) พบว่า วิธีกำหนดคะแนนจุดตัดทั้งสองให้คะแนนจุดตัดที่ใกล้เคียงกัน และผู้ตัดสินมีความเห็นว่าวิธีของอิมพาราและเพลคมีความสะดวกในการปฏิบัติมากกว่าวิธีของแองกอฟแบบดั้งเดิม

เพลคและเคน (Plake and Kane, 1991) เปรียบเทียบวิธีการกำหนดระดับการผ่านต่ำสุดที่ยอมรับได้จากการประมาณข้อสอบแต่ละข้อจากการตัดสินของผู้เชี่ยวชาญวิธีที่ใช้คือ วิธีที่ 1 ให้น้ำหนักระดับการผ่านต่ำสุดที่ยอมรับได้สำหรับข้อสอบตามระดับความสามารถที่เป็นสัดส่วนตรงข้ามกับความแปรปรวนอย่างสุ่มของการประมาณระดับการผ่านต่ำสุดที่ยอมรับได้ วิธีที่ 2 ให้น้ำหนักการประมาณระดับการผ่านต่ำสุดที่ยอมรับได้ของข้อสอบเท่ากัน ซึ่งเป็นวิธีกำหนดระดับการผ่านต่ำสุดที่ยอมรับได้จากการประมาณของผู้เชี่ยวชาญแบบดั้งเดิม และวิธีที่ 3 ผสมผสานระหว่างการทำนายระดับการผ่านต่ำสุดที่ยอมรับได้จากผู้เชี่ยวชาญและสัดส่วนของความสามารถต่ำสุดของผู้สอบที่ตอบข้อสอบแต่ละข้อได้ถูก พบว่าทั้งสามวิธีประมาณค่าได้ถูกต้องตามระดับความสามารถ เมื่อใช้ RMSE เป็นดัชนีความคงที่ โดยทั้ง 3 วิธีจะให้ความคงที่สูงสุดเมื่อระดับความสามารถเข้าคู่กับความยากโดยเฉลี่ยของข้อสอบ และจำนวนผู้เชี่ยวชาญ 10 คนจะมีความคงที่สูงกว่าผู้เชี่ยวชาญ 5 คน แต่ข้อสอบจำนวน 50 ข้อและ 25 ข้อมีความคงที่ไม่แตกต่างกัน และเมื่อแยกออกเป็นความลำเอียงและส่วนเบี่ยงเบนมาตรฐาน พบว่า วิธีที่ 3 มีความลำเอียงน้อยกว่า วิธีที่ 1 และ 2 แต่ขนาดความแตกต่างของความลำเอียงใน 3 วิธีน้อยมากเมื่อใช้ข้อสอบ 50 ข้อ และผู้เชี่ยวชาญ 10 คน สำหรับส่วนเบี่ยงเบนมาตรฐานของการประมาณค่าทั้งสามวิธีมีค่าใกล้เคียงกัน จากข้อสรุปแสดงว่า วิธีการกำหนดระดับการผ่านต่ำสุดที่ยอมรับได้ทั้ง 3 วิธีให้ข้อสรุปที่ใกล้เคียงกัน และวิธีที่ 2 ซึ่งใช้ผลรวมของค่าเฉลี่ยระดับการผ่านต่ำสุดที่ยอมรับได้ของข้อสอบเป็นคะแนนจุดตัดเป็นวิธีที่ง่ายกว่าอีก 2 วิธี

แพทริเซีย (Patricia, 1991) ศึกษาความเหมาะสมของการกำหนดจุดตัดด้วยวิธีของแองกอฟโดยใช้ผู้เชี่ยวชาญเฉพาะวิชา และได้รับการฝึกเกี่ยวกับการกำหนดคะแนนจุดตัด พบว่า การกำหนดคะแนนจุดตัดโดยใช้ดุลยพินิจของผู้เชี่ยวชาญไม่ได้ตัดสินโดยปราศจากเหตุผล และคะแนนจุดตัดที่ได้ขึ้นอยู่กับคุณภาพของผู้เชี่ยวชาญ และข้อมูลสารสนเทศประกอบการตัดสินที่เพียงพอ

นอร์ซินีและคณะ (Norcini, et al., 1991) ศึกษาผลของจำนวนผู้เชี่ยวชาญ และข้อสอบร่วมที่มีต่อความเท่ากันของคะแนนจุดตัดจากการตัดสินของผู้เชี่ยวชาญ พบว่า ผู้เชี่ยวชาญจำนวน 5 คน และข้อสอบร่วม 25 ข้อจะมีความคลาดเคลื่อนลดลง

นอร์ซินี (Norcini, 1990) ทดสอบความเท่ากันของการกำหนดคะแนนจุดตัดเพื่อตัดสินให้ผู้เรียนผ่านหรือตก พบว่า ความเท่ากันของคะแนนจุดตัดที่ตัดสินโดยผู้เชี่ยวชาญใกล้เคียงกับเกณฑ์มากกว่ามาตรฐานที่ได้จากการปฏิบัติการสอบของผู้สอบ เมื่อจำนวนของผู้สอบที่ใช้ในการเทียบปรับมีน้อย

นั่นคือ คะแนนจุดตัดจากการตัดสินของผู้เชี่ยวชาญให้ข้อสรุปที่ถูกต้องมากกว่าการเทียบปรับคะแนนเมื่อจำนวนผู้สอบมีน้อย

สเปรย์ และเวลค์ (Spray and Welch, 1990) ตรวจสอบการประมาณค่าเกี่ยวกับความคงที่ของการแบ่งกลุ่มผู้สอบเป็นกลุ่มผู้รอบรู้เมื่อความน่าจะเป็นเกี่ยวกับการตอบสนองข้อสอบได้ถูกต้องมีการแปรเปลี่ยนด้วยวิธีที่เสนอโดย SubKoviak (1976) พบว่าเมื่อความน่าจะเป็นเกี่ยวกับการตอบสนองข้อสอบได้ถูกต้องมีการแปรเปลี่ยน ไม่มีผลกระทบต่อการประมาณสัดส่วนความคงที่ของการแบ่งกลุ่ม

แวน เดอร์ ลินเดน (Van Der Linden, 1982) ใช้ทฤษฎีคุณลักษณะภายในสำหรับตัดสินความไม่สอดคล้องในการตัดสินของผู้ตัดสิน ในการกำหนดมาตรฐานด้วยวิธีของแองกอฟและนีเดลสกี พบว่า วิธีของนีเดลสกีมีความคลาดเคลื่อนมากกว่าวิธีของแองกอฟ

โคป (Cope, 1987) ให้ผู้เชี่ยวชาญประมาณความน่าจะเป็นในการทำแบบสอบได้ถูกต้องด้วยวิธีของแองกอฟ โดยแบ่งกลุ่มผู้เชี่ยวชาญเป็นให้ประมาณโดยไม่ได้รับสารสนเทศจากข้อสอบและได้รับสารสนเทศเกี่ยวกับค่าเฉลี่ยในการตัดสินและค่าความยากจากกลุ่มผู้สอบที่มีคะแนนใกล้คะแนนจุดตัดครั้งก่อน พบว่า ค่าเฉลี่ยของการประมาณหรือคะแนนจุดตัดระหว่างการไม่ได้รับสารสนเทศจากข้อสอบและได้รับสารสนเทศจากข้อสอบของผู้เชี่ยวชาญ แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 มีความสอดคล้องในการตัดสิน (Intrajudges Consistency) สูงสำหรับการประมาณของผู้เชี่ยวชาญที่ได้รับสารสนเทศจากข้อสอบ และค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างการประมาณของผู้เชี่ยวชาญครั้งที่ 1 และ 2 มีค่าต่ำ

นอร์ซินี และคณะ (Norcini, et. al., 1987) ตรวจสอบความแปรผันในการกำหนดคะแนนจุดตัดด้วยวิธีของแองกอฟที่มีผลต่อข้อสรุปของความสอดคล้อง โดยใช้กระบวนการกลุ่ม พบว่า การตัดสินด้วยกระบวนการกลุ่มเพิ่มประสิทธิภาพในการกำหนดคะแนนจุดตัดด้วยวิธีของแองกอฟ

เคน (Kane, 1987) ใช้ทฤษฎีการตอบสนองข้อสอบในการวิเคราะห์ผลการตัดสินคะแนนจุดตัดด้วยวิธีของแองกอฟ และเปรียบเทียบวิธีวิเคราะห์ 3 วิธี ได้แก่ วิธีการวิเคราะห์แบบดั้งเดิม วิธีวิเคราะห์ตามทฤษฎีการสรุปอ้างอิง และวิธีวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ พบว่า วิธีการวิเคราะห์แบบดั้งเดิมไม่ใช่วิธีที่ดีที่สุด และได้สนับสนุนให้ใช้วิธีที่วิเคราะห์ตามทฤษฎีการสรุปอ้างอิง

เฟรดแมนและโฮ (Friedman and Ho, 1990) ให้ผู้เชี่ยวชาญจำนวน 11 คน จาก 11 รัฐในอเมริกา ตัดสินข้อสอบจำนวน 65 ข้อ หลังจากนั้นนำผลการตัดสินที่ได้มาให้ผู้เชี่ยวชาญกันอภิปราย ก่อนที่จะให้ผู้เชี่ยวชาญตัดสินข้อสอบอีกครั้ง พบว่า สามารถเพิ่มความสอดคล้องในการตัดสินระหว่างกลุ่มผู้เชี่ยวชาญ (Interjudges Consistency) และความสอดคล้องในการตัดสินของผู้เชี่ยวชาญแต่ละคน (Intrajudges Consistency) ด้วย