



บทที่ ๒

วรรณคดีที่เกี่ยวข้อง

เนื่องจากงานวิจัยนี้เกี่ยวข้องกับแนวคิดหลายอย่าง เช่น การให้คะแนน สถิติและ
ทฤษฎีการวัดผลที่เกี่ยวข้อง และความรู้เกี่ยวกับการอ่านเข้าใจความ ดังนั้น ในบทนี้ผู้วิจัย
จึงขอแยกกล่าวถึงเรื่องต่าง ๆ ดังกล่าวแล้วเป็น ๔ ตอน ดังนี้

ตอนที่ ๑ ทฤษฎีการให้คะแนนรายชื่อของแบบสอบและงานวิจัยเกี่ยวกับการให้
คะแนน

ตอนที่ ๒ มโนทัศน์และทฤษฎีการสอบที่เกี่ยวข้อง

ตอนที่ ๓ งานวิจัยที่เกี่ยวกับตัวแปรที่มีผลต่อค่าความตรงและความเที่ยง

ตอนที่ ๔ มโนทัศน์และงานวิจัยเกี่ยวกับการอ่านเข้าใจความ

ตอนที่ ๑ ทฤษฎีการให้คะแนนรายชื่อของแบบสอบและงานวิจัยเกี่ยวกับการให้คะแนน

๑. ทฤษฎีการให้คะแนนรายชื่อของแบบสอบ

วิธีการกำหนดค่าน้ำหนักคะแนนรายชื่อให้แบบสอบฉบับหนึ่ง ตัวแปรของ
คะแนนรวม ได้แก่ คะแนนของข้อสอบรายชื่อ และการให้คะแนนก็เหมือนกับการคิดค่าน้ำหนัก
คะแนนของแบบสอบหลายฉบับที่นำมารวมกัน (Wang and Stanley, ๑๙๗๐:๖๗๒; Stanley
and Wang, ๑๙๗๐:๓๐) ซึ่งมีแนวคิดดังนี้

ถ้ากำหนดให้

$X_1, X_2 \dots X_n =$ คะแนนของข้อสอบแต่ละข้อ

$W_1, W_2 \dots W_n =$ ค่าน้ำหนักคะแนนของข้อสอบแต่ละข้อ

ดังนั้น

$$\begin{aligned} \text{Var}(X_1 + X_2 + \dots + X_n) &= \text{Var}(X_1) + \text{Var}(X_2) + \dots & [1] \\ &+ \text{Var}(X_n) + 2\text{Cov}(X_1 X_2) + 2\text{Cov}(X_1 X_3) + \dots + \\ &+ 2\text{Cov}(X_{n-1} X_n). \end{aligned}$$

เมื่อเมตริกมีขนาด $n \times n$ แล้ว ความแปรปรวนที่ข้อสอบข้อที่ i (C_i) มีส่วนในความแปรปรวนทั้งหมด มีดังนี้

$$\begin{aligned} C_i &= \text{Cov}(X_i X_1) + \text{Cov}(X_i X_2) + \dots + \text{Cov}(X_i X_{i-1}) + & [2] \\ &\text{Var}(X_i) + \text{Cov}(X_i X_{i+1}) + \dots + \text{Cov}(X_i X_n). \end{aligned}$$

ในกรณีที่ข้อสอบมีค่าน้ำหนักคะแนน ดังนั้น ความแปรปรวนของคะแนนรวม

เป็นดังนี้

$$\begin{aligned} \text{Var}(W_1 X_1 + W_2 X_2 + \dots + W_n X_n) &= W_1^2 \text{Var}(X_1) + & [3] \\ &W_2^2 \text{Var}(X_2) + \dots + W_n^2 \text{Var}(X_n) + 2W_1 W_2 \text{Cov}(X_1 X_2) + \\ &2W_1 W_3 \text{Cov}(X_1 X_3) + \dots + 2W_{n-1} W_n \text{Cov}(X_{n-1} X_n). \end{aligned}$$

และความแปรปรวนของข้อสอบรายข้อที่มีส่วนร่วมในความแปรปรวนทั้งหมดของคะแนนรวม

เป็นดังนี้

$$\begin{aligned} C_i &= W_i W_1 \text{Cov}(X_i X_1) + W_i W_2 \text{Cov}(X_i X_2) + \dots & [4] \\ &+ W_i W_{i-1} \text{Cov}(X_i X_{i-1}) + W_i^2 \text{Var}(X_i) + W_i W_{i+1} \text{Cov}(X_i X_{i+1}) + \dots \\ &+ W_i W_n \text{Cov}(X_i X_n). \end{aligned}$$

แต่เนื่องจาก $\text{Cov}(X_i X_1) = \rho_{i1} \sigma_i \sigma_1$ ดังนั้น

$$\begin{aligned} C_i &= W_i W_1 \rho_{i1} \sigma_i \sigma_1 + W_i W_2 \rho_{i2} \sigma_i \sigma_2 + \dots + & [5] \\ &W_i W_{i-1} \rho_{i,i-1} \sigma_i \sigma_{i-1} + W_i^2 \sigma_i^2 + W_i W_{i+1} \rho_{i,i+1} \sigma_i \sigma_{i+1} + \dots \\ &+ W_i W_n \rho_{in} \sigma_i \sigma_n. \end{aligned}$$

จากสมการที่ ๕ จะเห็นได้ว่า ข้อสอบแต่ละข้อมีส่วนร่วมในความแปรปรวนร่วมของคะแนนรวมทั้งหมดมากน้อยเพียงใดขึ้นอยู่กับปัจจัยต่อไปนี้

๑. คำน้้ำหนักคะแนนของข้อสอบแต่ละข้อที่กำหนดขึ้น
๒. ค่าความแปรปรวนของคะแนนของข้อสอบ
๓. ความสัมพันธ์ระหว่างข้อสอบนั้น กับข้อสอบต่าง ๆ ที่เหลือของแบบสอบ
๔. ค่าเบี่ยงเบนมาตรฐานของข้อสอบข้ออื่น ๆ ของแบบสอบ

แต่ในกรณีที่คะแนนของข้อสอบแต่ละข้อเป็นคะแนนมาตรฐาน คือ หารทุก ๆ ค่าด้วยค่าเบี่ยงเบนมาตรฐานของข้อสอบแต่ละข้อ (σ_i) สมการที่ ๕ จะมีลักษณะดังนี้

$$C_i = w_i w_1 \rho_{i1} + w_i w_2 \rho_{i2} + \dots + w_i w_{i-1} \rho_{i,i-1} + w_i^2 + w_i w_{i+1} \rho_{i,i+1} + \dots + w_i w_n \rho_{in} \quad [6]$$

ดังนั้น ในกรณีที่ข้อสอบแต่ละข้อเป็นคะแนนมาตรฐาน ค่าความแปรปรวนร่วมของข้อสอบแต่ละข้อจะมากน้อยเพียงใดขึ้นอยู่กับปัจจัย ๒ อย่าง คือ

๑. คำน้้ำหนักคะแนนของข้อสอบข้อนั้น ๆ ที่กำหนดขึ้น
๒. ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างข้อสอบแต่ละข้อ

ปกติแล้ว ถ้าคะแนนของข้อสอบแต่ละข้อเป็นคะแนนมาตรฐาน คะแนนของข้อสอบนั้นโดยประมาณแล้วเท่ากับ ๑ + ผลบวกของค่าสหสัมพันธ์ระหว่างข้อสอบแต่ละข้อ กล่าวคือ

$$C_i = 1 + \sum_{j=1}^n \rho_{ij} \quad [7]$$

และ $C_i = 1 + (n-1)\bar{\rho}_{ij} \quad [8]$

เมื่อ $\bar{\rho}_{ij} = \bar{x}$ ของค่าสหสัมพันธ์ของข้อสอบที่เหลือ (n-1)

ดังนั้น จึงกล่าวได้ว่าปกติแล้ว คำน้่านักคะแนนรายข้อที่กำหนดให้มีค่าไม่เท่ากับค่าที่กำหนดให้โดยตรง เว้นแต่จะทำให้คะแนนเป็นคะแนนมาตรฐาน และค่าความสัมพันธ์ซึ่งกันและกันของข้อสอบมีค่าเท่า ๆ กัน ตามทฤษฎีแล้ว การกำหนดค่าน้่านักคะแนนรายข้อทำให้แบบสอบมีความแปรปรวนของคะแนนมากกว่าการคิดคะแนนวิธีประเพณีนิยม ซึ่งจะสังเกตได้จากสมการที่ ๕ และ ๖ ดังนั้น จึงเชื่อว่าการคิดค่าน้่านักคะแนนเป็นวิธีหนึ่งที่จะทำให้แบบสอบมีค่าความเที่ยงและความตรงสูงกว่าค่าความเที่ยงและความตรงที่คิดจากคะแนนวิธีประเพณีนิยมได้ (Wang and Stanley, ๑๙๗๐:๖๖๓) ดังนั้น จึงมีการศึกษาเรื่องดังกล่าวนี้เรื่อยมา

๒. งานวิจัยเกี่ยวกับการให้คะแนน

ในการวัดผลนั้น วิธีที่นิยมกันทั่วไปในการให้คะแนน ได้แก่ การให้คะแนนข้อที่ถูกได้ ๑ คะแนน และข้อที่ผิดได้ ๐ คะแนน แต่เพื่อต้องการให้ค่าความตรงและค่าความเที่ยงสูงขึ้น จึงมีผู้พยายามคิดค้นหาวิธีที่จะให้คะแนนอีกหลายวิธี และได้มีงานวิจัยอีกเป็นจำนวนมากที่แสดงว่า ผลของความพยายามนี้มีทั้งประสบผลสำเร็จและล้มเหลว ซึ่งจะได้กล่าวต่อไป

ก. การให้คะแนนตัวเลือกแตกต่างกัน

การให้คะแนนตัวเลือกมีหลายวิธี แต่ละวิธีมีผลต่อความตรงและความเที่ยงของแบบสอบแตกต่างกัน กล่าวคือ

๑. วิธีอาศัยความเห็นของผู้เชี่ยวชาญ

จากการศึกษาของผู้วิจัยพบว่า การให้คะแนนตัวเลือกของข้อสอบโดยผู้เชี่ยวชาญ มี ๒ วิธีคือ ๑) อาศัยความเห็นของผู้เชี่ยวชาญว่าตัวเลือกแต่ละตัวมีความถูกต้องมากน้อยเพียงใด และ ๒) ใช้ค่าเฉลี่ยของลำดับที่ของความถูกต้องของตัวเลือกแต่ละตัวที่ตัดสินโดยผู้เชี่ยวชาญ จากการศึกษาของ Nedelsky ในปี ๑๙๔๔ (อ้างจาก Wang and Stanley, ๑๙๗๐:๖๗๑) ที่ใช้นักศึกษาที่สอบผ่านการสอบรวบรวมแล้ว ๓๐๖ คน

เป็นกลุ่มตัวอย่าง โดยกำหนดค่าน้ำหนักคะแนนโดยวิธีที่ ๑ ปรากฏว่า แบบสอบแบบเลือกตอบ มีค่าความเที่ยงสูงกว่าค่าความเที่ยงที่คำนวณจากคะแนนวิธีประเพณีนิยม และในปี ๑๙๖๘ Jacobs และ Vanderventer (อ้างจาก Wang and Stanley, ๑๙๗๐:๖๕๓) ได้ใช้วิธีการให้คะแนนวิธีเดียวกัน ปรากฏว่า การให้คะแนนดังกล่าวทำให้แบบสอบแบบเลือกตอบ มีค่าความเที่ยงจากการสอบซ้ำ ค่าความตรงเชิงพยากรณ์ และค่าความตรงร่วมสมัย (Concurrent Validity) มีค่าสูงขึ้นกว่าค่าความเที่ยงและความตรงชนิดเดียวกัน ที่คำนวณจากคะแนนที่ให้โดยวิธีประเพณีนิยม

ส่วนการให้คะแนนวิธีที่ ๒ นั้น Hambleton, Roberts และ Traub ได้ทำการศึกษาในปี ๑๙๗๐ (อ้างจาก Wang and Stanley, ๑๙๗๐:๖๕๓) ปรากฏว่า การให้คะแนนวิธีดังกล่าวทำให้แบบสอบแบบเลือกตอบมีค่าความเที่ยงและค่าความตรงเชิงพยากรณ์สูงเกินกว่าค่าความเที่ยงและค่าความตรงที่คำนวณจากคะแนนที่ให้โดยวิธีประเพณีนิยม แต่ไม่มีนัยสำคัญทางสถิติ

๒. วิธีใช้ค่าความสัมพันธ์ระหว่างตัวเลือกกับสิ่งที่มีงทดสอบ

การให้คะแนนโดยอาศัยแนวคิดเกี่ยวกับความสัมพันธ์ระหว่างตัวเลือก กับสิ่งที่มีงทดสอบ เท่าที่พบมี ๓ วิธีคือ ๑) ใช้ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวเลือกกับคะแนนรวม ๒) ใช้ค่ามาตรวัดที่คำนวณจากวิธี Guttman's Scale Analysis และ ๓) ใช้ค่ามาตรวัดที่คำนวณจากวิธีการเปรียบเทียบรายคู่ของ Thurstone (Thurstone's Paired Comparison) จากการศึกษาวรรณคดีที่เกี่ยวข้องพบว่า Davis และ Fifer ได้ใช้ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวเลือกแต่ละตัวกับคะแนนรวม เป็นคะแนนแบบสอบคู่นานวิชาคณิตศาสตร์ ซึ่งยาว ๔๕ ข้อ ในปี ๑๙๕๔ (อ้างจาก Wang and Stanley, ๑๙๗๐:๖๕๒) ผลปรากฏว่า การให้คะแนนวิธีดังกล่าวทำให้แบบสอบมีค่าความเที่ยงสูงมากกว่าค่าความเที่ยงที่คำนวณจากคะแนนที่ให้โดยวิธีประเพณีนิยมอย่างมีนัยสำคัญ คือ เพิ่มขึ้นจาก ๐.๖๘ เป็น ๐.๗๖ ซึ่งถ้าจะทำให้แบบสอบนั้นมีค่าความเที่ยงเท่ากับค่าที่เพิ่มขึ้น เมื่อให้คะแนนวิธีประเพณีนิยม จะต้องเพิ่มข้อสอบอีก ๔๕-๖๗ ข้อ

ส่วนการให้คะแนนโดยใช้ค่ามาตรวัดจากวิธี Guttman's Scale Analysis นั้น Hendrickson ได้ทำการศึกษาในกลุ่มตัวอย่างที่เป็นเพศชาย ๕,๐๐๐ คน และเพศหญิง ๕,๐๐๐ คน ที่สอบ Scholastic Aptitude Test ในปี ๑๙๗๐ และ Raffeld ได้ทำการศึกษาในกลุ่มตัวอย่างที่เป็นนักเรียน Grade ๘ จำนวน ๑,๓๕๔ คน ที่สอบวิชาการอ่านเข้าใจความ ศัพท์ ไวยากรณ์ การสะกดคำ และคณิตศาสตร์ (อ้างจาก Wang and Stanley, ๑๙๗๐:๖๔๓; Raffeld, ๑๙๗๕:๒๘๕) จากการศึกษา Hendrickson พบว่า ความเที่ยงของแบบสอบสูงเกินกว่าค่าความเที่ยงที่คำนวณจากคะแนนที่ให้โดยวิธีประเพณีนิยมอย่างมีนัยสำคัญ ซึ่งถ้าให้คะแนนวิธีประเพณีนิยม และต้องการให้แบบสอบนั้นมีค่าความเที่ยงสูงเท่ากับค่าความเที่ยงที่เพิ่มขึ้น จะต้องเพิ่มข้อสอบอีกประมาณ ๔๔% ในทำนองเดียวกัน Raffeld พบว่า ความเที่ยงและความตรงเชิงพยากรณ์ของแบบสอบที่ให้คะแนนโดยใช้มาตรวัดวิธี Guttman's Scale Analysis มีค่าสูงกว่าค่าความเที่ยงและความตรงเชิงพยากรณ์ที่คำนวณจากคะแนนที่ให้โดยวิธีประเพณีนิยมอย่างมีนัยสำคัญ เมื่อใช้คะแนนปลายปีเป็นเกณฑ์สำหรับการพยากรณ์

แต่อย่างไรก็ตาม เมื่อ Patnaik และ Traub (Patnaik and Traub, ๑๙๗๓:๒๘๑) ทำการศึกษาผลของการให้คะแนนต่อค่าความเที่ยงและความตรงเชิงพยากรณ์ในปี ๑๙๗๓ กับนักศึกษาในระดับบัณฑิตศึกษา จำนวน ๖๑ คน โดยใช้แบบสอบวัดเชาวน์ปัญญาที่เป็นแบบสอบแบบเลือกตอบ ชนิดมี ๕ ตัวเลือก และใช้มาตรวัดจากวิธีการเปรียบเทียบรายคู่ของ Thurstone เป็นคะแนนของตัวเลือกแต่ละตัว ผลปรากฏว่า การให้คะแนนดังกล่าวทำให้แบบสอบมีค่าความเที่ยงแบบแบ่งครึ่งสูงกว่าค่าความเที่ยงที่คำนวณจากคะแนนที่ให้โดยวิธีประเพณีนิยมอย่างมีนัยสำคัญ ($p < .๐๕$) แต่มีค่าความตรงเชิงพยากรณ์ที่ใช้ผล การสอบสัมฤทธิ์ผลวิชาต่าง ๆ เป็นเกณฑ์ต่ำกว่าค่าความตรง ชนิดดังกล่าวอย่างมีนัยสำคัญ ($p < .๐๕$) การที่ค่าความเที่ยงของแบบสอบเพิ่มขึ้น แต่ค่าความตรงเชิงพยากรณ์ลดลง จากค่าความตรงเชิงพยากรณ์ที่คำนวณจากคะแนนที่ให้โดยวิธีประเพณีนิยมนั้น Patnaik และ Traub อธิบายว่า เป็นเพราะว่าการให้คะแนนที่ไม่ใช้วิธีประเพณีนิยมทำให้แบบสอบมีค่าสหสัมพันธ์ระหว่างข้อสอบเพิ่มขึ้น เพราะคะแนนมีการกระจายเพิ่มขึ้นมากกว่าเดิม

ด้วยเหตุนี้จึงทำให้คะแนนจากแบบสอบมีความสัมพันธ์กับเกณฑ์ต่ำ เช่นเดียวกับกรณีที่ว่า
สัมประสิทธิ์สหสัมพันธ์พหุคูณของตัวแปรที่มีความสัมพันธ์ซึ่งกันและกันสูง มีความสัมพันธ์กับ เกณฑ์
ต่ำ (Patnaik and Traub, ๑๙๗๗:๒๘๕)

๓. วิธีอาศัยความเชื่อมั่นของผู้สอบต่อตัวเลือกของแบบสอบ

วิธีการให้คะแนนโดยอาศัยความเชื่อมั่นของผู้สอบต่อตัวเลือกของแบบ
สอบ อาจแบ่งออกได้ ๓ วิธีคือ ๑) วิธีคิดคะแนนตามความมั่นใจในการตอบ ๒) วิธีคิด
คะแนนโดยหักค่าการเดา และ ๓) วิธีคิดคะแนนโดยวิธีการกำจัดตัวเลือก สำหรับการคิด
คะแนนวิธีที่ ๑ นั้น แต่ละข้อถ้าผู้สอบตอบถูกในการเลือกตอบครั้งแรกจะได้ ๓ คะแนน
ถ้าเลือก ๒ ครั้ง ตอบถูกได้ ๒ คะแนน ถ้าเลือก ๓ ครั้ง ตอบถูกได้ ๑ คะแนน และถ้า
เลือก ๔ ครั้ง ตอบถูกจะได้ ๐ คะแนน (ข้อสอบมี ๔ ตัวเลือก) วิธีที่ ๒ ให้คะแนนโดย
อาศัยสูตรการคำนวณที่หักค่าการเดา และวิธีที่ ๓ นั้น ผู้สอบเลือกตอบหลาย ๆ ตัวเลือกใน
แต่ละข้อ ถ้าตอบถูกจะได้ ๑ คะแนน แต่ถ้าตอบผิดจะได้ $-\frac{1}{3}$ คะแนน (ข้อสอบมี ๔ ตัวเลือก)

จากการศึกษาของ Diamond (Diamond, ๑๙๗๕:๑๒๙) ในปี ๑๙๗๕
เกี่ยวกับการให้คะแนนทั้ง ๓ วิธี กับกลุ่มตัวอย่าง ๘๔ คน ในระดับบัณฑิตศึกษา ด้วยแบบสอบ
วิชาวัดผลที่ค่อนข้างง่าย และมีอำนาจจำแนกต่ำ ผลปรากฏว่า แบบสอบที่ให้คะแนนทั้ง ๓ วิธี
ดังกล่าวมีค่าความเที่ยงแบบ Flanagan สูงกว่าค่าความเที่ยงที่คำนวณจากคะแนนที่ให้โดย
วิธีประเพณีนิยม แต่ว่าค่าความตรงเชิงพยากรณ์ของแบบสอบที่ใช้ผลการสอบประจำภาค
เป็นเกณฑ์ ไม่แตกต่างจากค่าความตรงเชิงพยากรณ์ที่คำนวณจากคะแนนที่ให้โดยวิธีประเพณี
นิยม ผลการศึกษาครั้งนี้มีความคล้ายคลึงกับผลการศึกษาของ Hanna ในปีเดียวกัน
(Hanna, ๑๙๗๕:๑๔๕) ซึ่งได้ให้คะแนนตัวเลือกของแบบสอบแบบเลือกตอบชนิด ๔ ตัวเลือก
๒ ฉบับ ซึ่งยาว ๑๐ ข้อ และ ๓๐ ข้อ และใช้กับกลุ่มตัวอย่าง ๘๕ คน โดยวิธีให้คะแนนตาม
ความมั่นใจในการตอบ ผลการศึกษาพบว่า แบบสอบที่ให้คะแนนด้วยวิธีดังกล่าว มีค่าความ
เที่ยงแบบแบ่งครึ่งสูงกว่าค่าความเที่ยงที่คำนวณจากคะแนนที่ให้โดยวิธีประเพณีนิยมเล็กน้อย
แต่ค่าความตรงเชิงพยากรณ์ซึ่งใช้คะแนนสอบการตีความ และการเขียนรายงานเป็น เกณฑ์กลับ

มีค่าต่ำกว่าค่าความตรงเชิงพยากรณ์ที่ได้จากคะแนนที่ให้โดยวิธีประเพณีนิยม ดังนั้น Hanna จึงสรุปว่า การให้คะแนนวิธีประเพณีนิยมให้ผลการสอบดีกว่าวิธีการให้คะแนนตามความมั่นใจในการตอบ เพราะการสอบนั้นควรคำนึงถึงความตรงของแบบสอบมากกว่าความเที่ยง

นอกจากนี้ ในปี ๑๙๗๕ Hakstian และ Kansup (Hakstain and Kansup, ๑๙๗๕:๒๓๑) ยังทำการศึกษาผลการให้คะแนนโดยวิธีให้คะแนนตามความมั่นใจในการตอบ และวิธีให้คะแนนโดยวิธีการกำจัดตัวเลือก โดยใช้นักเรียนระดับ grade ๕ จำนวน ๑,๐๒๓ คน เป็นกลุ่มตัวอย่าง ผลการศึกษาพบว่า สอดคล้องกับผลการศึกษาของ Diamond และ Hanna กล่าวคือ การให้คะแนนทั้ง ๒ วิธี ทำให้แบบสอบมีค่าความเที่ยงสูงขึ้นเล็กน้อย แต่ให้ค่าความตรงเชิงพยากรณ์ต่ำกว่าค่าความตรงที่คำนวณจากคะแนนที่ให้โดยวิธีประเพณีนิยม

อนึ่ง เพื่อเป็นการตรวจสอบผลการศึกษาที่ทำไว้ในปี ๑๙๗๕ Hanna ได้ทำการศึกษาเกี่ยวกับผลการให้คะแนน โดยวิธีให้คะแนนตามความมั่นใจในการตอบอีกในปี ๑๙๗๗ ครั้งนี้ Hanna (Hanna, ๑๙๗๗:๖๔๓) ใช้แบบสอบแบบเลือกตอบ ยาว ๑๘ ข้อ ๒ ฉบับ คือ วิชาวิทยาศาสตร์และคณิตศาสตร์ และใช้นักเรียน grade ๕ และ ๖ จำนวน ๑,๓๘๑ คน เป็นตัวอย่าง ผลการศึกษาปรากฏว่า มีความสอดคล้องกับผลการศึกษาในปี ๑๙๗๕ คือ แบบสอบมีค่าความเที่ยงสูงกว่าค่าความเที่ยงที่คำนวณจากคะแนนที่ให้โดยวิธีประเพณีนิยมอย่างมีนัยสำคัญ แต่ให้ค่าความตรงไม่แตกต่างจากค่าความตรงที่คำนวณจากคะแนนที่ให้โดยวิธีประเพณีนิยม

ข. การให้ค่านำหนักคะแนนข้อสอบแตกต่างกัน

นอกจากให้คะแนนตัวเลือกของข้อสอบแต่ละข้อแตกต่างกันแล้ว นักทดสอบยังสนใจที่จะให้คะแนนเฉพาะข้อสอบแตกต่างกันด้วย เพื่อศึกษาว่าการให้คะแนนดังกล่าวนี้จะทำให้แบบสอบมีค่าความเที่ยงและความตรงสูงกว่าค่าความเที่ยงและความตรงที่ได้จากคะแนนที่ให้โดยวิธีประเพณีนิยมหรือไม่

จากการศึกษาวรรณคดีที่เกี่ยวข้อง ผู้วิจัยพบว่า การให้คะแนนข้อสอบแตกต่างกัน มีหลายวิธีดังนี้

๑. วิธีอาศัยความเห็นของผู้เชี่ยวชาญ

การให้คะแนนข้อสอบแตกต่างกัน โดยอาศัยความเห็นของผู้เชี่ยวชาญที่เกี่ยวข้อง เป็นวิธีที่ Corey ได้นำมาศึกษาในปี ๑๙๓๐ และ Odell ได้ตรวจสอบผลการศึกษานั้นในปี ๑๙๓๑ (อ้างจาก Wang and Stanley, ๑๙๗๐:๖๘๘) Corey ให้ผู้เชี่ยวชาญทางจิตวิทยา ๖ คน พิจารณาให้คะแนนข้อสอบแตกต่างกันตามความสำคัญของสิ่งที่มุ่งทดสอบ โดยไม่ต้องปรึกษากัน ปรากฏว่าคะแนนของแบบสอบซึ่งมีคะแนนรายข้อแตกต่างกันทั้ง ๖ ชุด มีค่าสัมประสิทธิ์สหสัมพันธ์กับคะแนนที่ให้โดยวิธีประเพณีนิยม ระหว่าง .๘๓๖ ถึง .๘๖๐ ซึ่งนับว่าสูงมาก ผลของการศึกษาของ Odell ในปีต่อมาสอดคล้องกับผลการศึกษาของ Corey ทุกประการ ทำให้เขาสรุปว่า การให้คะแนนรายข้อแตกต่างกันไม่มีประโยชน์ในการสอบ

๒. วิธีใช้ค่าถดถอยพหุคูณ

ในปี ๑๙๔๒ Guilford, Lovell และ Williams (Guilford and Others, ๑๙๔๒:๑๕) ได้ทำการศึกษาผลของการให้คะแนนรายข้อแตกต่างกัน โดยอาศัยค่าประมาณของค่าถดถอยพหุคูณ ซึ่งคำนวณจากสูตรพิเศษที่คิดขึ้น คือ

$$w_i = \frac{\rho_{ic} \sigma_c}{\sigma_i}$$

ρ_{ic} = ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างข้อสอบกับเกณฑ์

σ_c = ค่าความเบี่ยงเบนมาตรฐานของ เกณฑ์

σ_i = ค่าความเบี่ยงเบนมาตรฐานของรายข้อ

ผลจากการให้คะแนนรายข้อแตกต่างกันด้วยสูตรดังกล่าวแล้ว กับแบบสอบแบบเลือกตอบ ๓ ฉบับ ที่มีความยาว ๒๐ ข้อ ๕๐ ข้อ และ ๑๐๐ ข้อ Guilford และคณะพบว่า ความเที่ยงแบบแบ่งครึ่ง และความตรงเชิงพยากรณ์ของแบบสอบ ซึ่งใช้คะแนนแบบสอบสัมฤทธิ์ผลยาว ๑๐๐ ข้อ เป็นเกณฑ์ มีค่าไม่แตกต่างจากค่าความเที่ยง และค่าความตรงเชิงพยากรณ์ที่คำนวณจากคะแนนที่ให้โดยวิธีประเพณีนิยมอย่างมีนัยสำคัญ ดังนั้น Guilford และคณะจึงสรุปว่า การให้คะแนนรายข้อแตกต่างกันได้ผลไม่คุ้มค่ากับความยุ่งยากในการคำนวณ

นอกจากนี้แล้ว ยังมีนักทดสอบอีกจำนวนมากที่ให้คะแนนรายข้อต่างกันด้วยวิธีต่าง ๆ แต่ปรากฏว่า ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนที่ได้จากการให้คะแนนรายข้อต่างกัน กับคะแนนที่ให้โดยวิธีประเพณีนิยม มีความสัมพันธ์กันสูงมาก เช่น Douglas และ Spencer (๑๙๒๓) พบว่า ค่าดังกล่าวสัมพันธ์กันระหว่าง .๙๘-.๙๙ Holzinger (๑๙๒๓) West (๑๙๒๔) Peatman (๑๙๓๐) Ruch และ Meyer (๑๙๓๑) Pothoff และ Barnett (๑๙๓๒) และคนอื่น ๆ อีกต่างก็พบว่า ความสัมพันธ์ระหว่างคะแนนดังกล่าวมีความสัมพันธ์กันสูงระหว่าง .๙๓-.๙๔ (อ้างจาก Wang and Stanley, ๑๙๗๐:๖๘๘)

ดังนั้น นักทดสอบจำนวนมากจึงมีความเห็นว่า การให้คะแนนรายข้อของแบบสอบต่างกันไม่มีประโยชน์ เพราะว่าคะแนนที่ได้มีความสัมพันธ์กับคะแนนที่ให้โดยวิธีประเพณีนิยมมาก ด้วยเหตุนี้ Wang and Stanley (Wang and Stanley, ๑๙๗๐:๖๘๘) ผู้ซึ่งศึกษาและรวบรวมวรรณคดีที่เกี่ยวกับการให้ค่านำหนักคะแนนแบบสอบวิธีต่าง ๆ จนกระทั่งถึงปี ๑๙๗๐ ได้สรุปว่า การให้คะแนนวิธีต่าง ๆ ไม่มีผลดีต่อคุณภาพของแบบสอบ หรือถ้ามีผลดีก็มีเพียงเล็กน้อยเท่านั้น เพราะว่าในทางปฏิบัติวิธีการให้คะแนนเหล่านั้นมักนำไปใช้ไม่ได้ แต่ก็ยังพอมีความหวังอีก ๒ วิธี ที่จะนำมาใช้ในการให้คะแนนข้อสอบ คือ

๑. วิธีให้คะแนนโดยอาศัยรูปแบบความสามารถแฝง
๒. วิธีให้คะแนนโดยอาศัยรูปแบบการถดถอยพหุคูณ

๓. วิธีให้คะแนนโดยอาศัยรูปแบบความสามารถแฝง

Lord (Lord, ๑๙๖๘:๔๘๔) ได้ทำการวิเคราะห์ผลการสอบวิชา Scholastic Aptitude Test โดยให้คะแนนของข้อสอบต่างกันตามแนวคิดของ Birnbaum โดยอาศัยรูปแบบโลจิสติกใช้พารามิเตอร์ ๓ ตัว ทั้งนี้โดยการกำหนดค่าน้ำหนักคะแนนของผู้สอบแต่ละคน และแต่ละข้อต่างกันตามความสามารถของแต่ละบุคคล ซึ่งจะทราบได้จากโค้งลักษณะของข้อสอบ (Item Characteristic Curve) แนวคิดดังกล่าวนี้ต่อมา Lord (Lord, ๑๙๗๗:๑๒๓) ได้ย้่าว่า เป็นแนวคิดที่สำคัญ กล่าวคือ เขาคิดว่า "ข้อสอบข้อที่ให้คะแนนมากสำหรับคนเก่ง ไม่จำเป็นต้องให้คะแนนมากสำหรับคนไม่เก่ง" และสูตรในการคำนวณหาค่าน้ำหนักคะแนนที่เหมาะสมแต่ละข้อ คือ

$$w_i = \frac{a_i}{1 - c_i} \cdot \frac{p_i^{-c_i}}{p_i}$$

เมื่อ a_i = ค่าอำนาจจำแนกของข้อสอบข้อที่ i

c_i = ค่าการเดาของข้อสอบข้อที่ i

p_i = ค่าความยากง่ายของข้อสอบข้อที่ i

ผลของการวิเคราะห์ผลการสอบ Lord พบว่า

๑. เมื่อความสามารถของผู้สอบมีมาก โค้งแสดงค่าน้ำหนักคะแนนจะขนานกับแกนนอน
๒. เมื่อความสามารถของผู้สอบต่ำ โค้งแสดงค่าน้ำหนักคะแนนจะขึ้นอยู่กับความยากของข้อสอบ

ดังนั้น เขาจึงสรุปว่า ในกรณีที่ผู้สอบมีความสามารถสูงมาก การให้คะแนนข้อสอบรายข้อวิธีประเพณีนิยมจะดีกว่า เพราะว่าค่าน้ำหนักคะแนนระดับสูง ๆ จะคงที่ แต่ถ้าผู้สอบที่มีความสามารถต่ำหรือปานกลาง การให้คะแนนข้อสอบต่างกันมีความถูกต้องมากกว่า

อย่างไรก็ตาม เนื่องจากผลการวิจัยเกี่ยวกับการให้คะแนนตัวเลือก หรือข้อสอบต่างกันได้ผลไม่แน่นอน พอที่จะสรุปรวมได้ด้วยความมั่นใจว่า การให้คะแนนดังกล่าวทำให้แบบสอบมีความเที่ยงและความตรงเพิ่มขึ้นได้จริง ในปี ๑๙๗๕ Lord (Lord, ๑๙๗๕: ๑๐) ศึกษาค้นคว้าที่มีชื่อเสียงมากคนหนึ่งในปัจจุบันนี้ ได้สรุปความคิดเห็นเกี่ยวกับเรื่องนี้ที่น่าสนใจมากไว้ดังต่อไปนี้

๑. การให้คะแนนวิธีประเพณีนิยมแล้ว ผู้สอบทำไม่ครบหมดทุกข้อ ผลของการสอบจะเป็นตัวทำนายได้ไม่ดีเท่ากับการให้คะแนนโดยวิธีให้คะแนนข้อสอบต่างกัน

๒. การให้คะแนนโดยวิธีให้คะแนนข้อสอบต่างกันแล้ว ทำให้แบบสอบมีค่าความเที่ยง หรือค่าความตรงเพิ่มขึ้นแม้เพียงเล็กน้อย เช่น .๕๐ เป็น .๕๑ หรือ .๖๐ เป็น .๖๐๓๓ ก็ถือว่าเป็นการพัฒนาแบบสอบที่คุ้มค่า เพราะสามารถลดจำนวนข้อสอบลงได้อย่างน้อยประมาณ $\frac{๑}{๔}$ จากจำนวนข้อสอบ

๓. การศึกษาเปรียบเทียบค่าความเที่ยงของแบบสอบ โดยการให้คะแนนข้อสอบต่างกันไม่มีประโยชน์ เว้นเสียแต่ว่าจะสามารถพิสูจน์ว่า วิธีที่นำมาเปรียบเทียบกันวัดคุณลักษณะ (traits) เดียวกันอย่างแท้จริง

ต่อมาในปี ๑๙๘๐ Lord (Lord, ๑๙๘๐:๗๔) ได้ศึกษาเปรียบเทียบผลของการให้คะแนนข้อสอบแต่ละข้อแตกต่างกันอีก โดยใช้แบบสอบทางภาษาชื่อ SCAT II Form 2A ซึ่งมี ๔ ตัวเลือก จำนวน ๔๐ ข้อ เป็นเครื่องมือ เขาพบว่า การให้คะแนนข้อสอบแต่ละข้อโดยวิธีใช้ค่าน้ำหนักคะแนนที่เหมาะสม (w) ให้สาระของการสอบมากที่สุด แต่สำหรับผู้สอบที่มีความสามารถในระดับต่ำ การให้คะแนนวิธีประเพณีนิยม ให้สาระดังกล่าว น้อยที่สุด ถ้าการให้คะแนนดังกล่าวโดยอาศัยรูปแบบโลจิสติกใช้พารามิเตอร์ ๓ ตัวแล้ว การให้คะแนนวิธีประเพณีนิยมให้สาระการสอบได้ดีกว่า ทั้งนี้เพราะรูปแบบดังกล่าวไม่ได้ใช้ค่าการเดาด้วย และในปี ๑๙๘๒ Hsu และ Choosak (Hsu and C. Khampalikit, ๑๙๘๒:๗) ได้เปรียบเทียบผลการให้คะแนนโดยวิธีใช้ค่าน้ำหนักคะแนนที่เหมาะสม (w) กับการให้คะแนนวิธีประเพณีนิยมแก่ข้อสอบแต่ละข้อ สำหรับวิชาภาษาจีนและวิชาประวัติศาสตร์ เขาพบว่า การให้คะแนน

วิธีแรกทำให้แบบสอบมีค่าอำนาจจำแนกและค่าความเที่ยงสูงกว่าค่าอำนาจจำแนก และค่าความเที่ยงที่คำนวณจากคะแนนที่ให้โดยวิธีประเพณีนิยม แต่ไม่ได้ศึกษาเกี่ยวกับผลที่เกิดขึ้นกับค่าความตรงของแบบสอบดังกล่าว

อนึ่ง สำหรับการให้คะแนนโดยวิธีอาศัยรูปแบบการถดถอยพหุคูณนั้น เป็นแนวคิดใหม่ที่มีผู้กำหนดคะแนนแก่ผู้สอบแต่ละคนแตกต่างกันตามความสามารถของแต่ละบุคคล รายละเอียดของวิธีนี้ยังไม่อาจหาได้ ส่วนการให้คะแนนรายข้อหรือแต่ละตัวเลือก โดยวิธีใช้ค่าถดถอยพหุคูณนั้น นักทดสอบใช้กันมานานแล้ว แต่ปรากฏว่ายังทำให้แบบสอบมีค่าความเที่ยงและความตรงเพิ่มขึ้นอย่างมีนัยสำคัญเฉพาะแบบสอบที่มีขนาดสั้นเท่านั้น (Wang and Stanley, ๑๙๗๐:๖๗๐)

ดังนั้น จะเห็นได้ว่าวิธีที่จะทำให้ค่าความตรงและค่าความเที่ยงเพิ่มขึ้นในการวัดคุณลักษณะเดียวกัน โดยวิธีการให้คะแนนรายข้อหรือแต่ละตัวเลือกต่างกัน จึงยังเป็นแนวคิดที่น่าสนใจสำหรับนักทดสอบทางการศึกษาต่อไป

ตอนที่ ๒ มโนทัศน์และทฤษฎีการสอบที่เกี่ยวข้อง

ก. ความตรงของแบบสอบอิงกลุ่ม

๑. ความหมายของความตรง

โดยทั่วไปนั้น ความตรงหมายถึง ความถูกต้องของสิ่งที่วัดที่แบบสอบนั้นต้องการวัด (Ebel, ๑๙๗๒:๔๐๕) แต่ความหมายเฉพาะของความตรงนั้นแตกต่างกันตามวิธีการคำนวณ และจุดมุ่งหมายของการวัด

๒. ชนิดต่าง ๆ ของความตรง

ความตรงมีหลายชนิด (Ebel, ๑๙๗๒:๔๓๖; Cronbach, ๑๙๕๕:๑๐๓) เช่น

ก. ความตรงเชิงพยากรณ์ (Predictive Validity)

ได้แก่ ความถูกต้องในการพยากรณ์พฤติกรรม หรือสิ่งที่ต้องการวัดที่จะเกิดขึ้นในอนาคตของแบบสอบ ซึ่งได้จากค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนของแบบสอบนั้น กับคะแนนการสอบของแบบสอบที่ใช้เป็นเกณฑ์

ข. ความตรงร่วมสมัย (Concurrent Validity) ได้แก่

ความถูกต้องในการวัดของแบบสอบฉบับหนึ่ง ที่มีความถูกต้อง เช่นเดียวกับผลการวัดของแบบสอบอีกฉบับหนึ่ง ที่มุ่งวัดสิ่งเดียวกันในปัจจุบัน ค่าความตรงนี้คำนวณได้จากค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนการสอบของแบบสอบทั้ง ๒ ฉบับ

ค. ความตรงเชิงภาวะสันนิฐาน (Construct Validity)

ได้แก่ ความถูกต้องตามภาวะสันนิฐานของโครงสร้างของสิ่งที่วัด ซึ่งมักได้จากการทดสอบโครงสร้างของสิ่งที่วัดโดยวิธีการทางสถิติวิธีใดวิธีหนึ่ง

ง. ความตรงตามนิยาม (Face Validity) ได้แก่ ความ

ถูกต้องของแบบสอบตามรูปแบบ หรือลักษณะของแบบสอบที่ควรจะเป็น เท่าที่สังเกตได้โดยตาเปล่า

จ. ความตรงเชิงเนื้อหา (Content Validity) ได้แก่

ความถูกต้องและเพียงพอของสิ่งที่มุ่งวัด

๓. ปัจจัยสำคัญที่มีผลต่อค่าความตรง

ปัจจัยสำคัญที่มีผลต่อค่าความตรง มี ๒ อย่าง (Gulliksen, ๑๙๕๘:

๘๘) คือ

๑. ความยาวของแบบสอบ แบบสอบที่ยาวยิ่งทำให้ความตรงของแบบสอบมีมากขึ้น

๒. การกระจายของคะแนนการสอบ แบบสอบที่ยิ่งมีคะแนนแตกต่างกันมาก ก็ยิ่งมีค่าความตรงมากขึ้น

๔. ความสัมพันธ์ระหว่างความตรงและความเที่ยงของแบบสอบ

แบบสอบที่มีความตรงจำเป็นอย่างยิ่งที่จะต้องมีความเที่ยงด้วย ดังนั้นค่าความเที่ยงจึงมีความจำเป็นต่อค่าความตรง แต่แบบสอบที่มีความเที่ยงสูง อาจมีค่าความตรงต่ำ เพราะแบบสอบนั้นอาจไม่ได้วัดสิ่งที่มุ่งทำการสอบก็ได้ (Ebel, ๑๙๗๒:๔๔๔) ตามทฤษฎีการสอบเดิม แบบสอบที่มีความยาวเพิ่มขึ้นด้วยข้อสอบที่วัดสิ่งเดียวกัน และมีคุณสมบัติเหมือนกับข้อสอบเดิมที่มีอยู่ จะทำให้ค่าความเที่ยงเพิ่มขึ้น และทำให้ค่าความตรงเพิ่มขึ้นด้วย แต่ค่าสูงสุดของความตรงมีได้ไม่เกินรากกำลังสองของค่าความเที่ยง (r_{tt}) ของแบบสอบนั้น (Magnuson, ๑๙๖๖:๑๕๐)

ข. ความเที่ยงของแบบสอบอิงกลุ่ม

๑. แนวคิดของค่าความเที่ยง

ความเที่ยงของแบบสอบอิงกลุ่ม มีความหมายได้หลายอย่างอย่างหนึ่งได้แก่ "ดัชนีของการทำนายค่าคะแนนของแบบสอบฉบับหนึ่งจากแบบสอบคู่ขนานอีกชุดหนึ่ง" (Ghiselli, ๑๙๖๔:๒๑๔) ในการกล่าวถึงค่าความเที่ยงนั้น มักจะบอกค่าความคลาดเคลื่อนมาตรฐานในการวัด (SE_m) ซึ่งในรูปของความแปรปรวน ได้แก่ รากกำลังสองของความแปรปรวนของความคลาดเคลื่อนสัมพันธ์ (relative error variance: $\sigma^2(\delta)$) นั้นเอง ความเที่ยงของแบบสอบอิงกลุ่มมีแนวคิดต่างกัน ๓ อย่าง (Ghiselli, ๑๙๖๔:๒๑๔-๒๕๐) คือ

ก. ความเที่ยงตามแนวคิดของทฤษฎีคะแนนจริง และคะแนนคลาดเคลื่อน (Theory of True and Error Scores)

แนวคิดนี้มีสมมุติฐานว่า ๑) คะแนนที่ปรากฏประกอบด้วยคะแนนจริง และคะแนนของความคลาดเคลื่อน ๒) ความคลาดเคลื่อนในการวัดนั้นไม่มีระบบแน่นอน และ ๓) การสอบแต่ละครั้งผู้รับการสอบมีสภาพเหมือนเดิมทุกครั้ง หากว่ากลุ่มตัวอย่างจำนวนหนึ่งทำการสอบแบบสอบคู่ขนานกันจำนวนอนันต์ครั้งแล้ว ค่าสหสัมพันธ์ระหว่างคะแนนของแบบสอบคู่ขนานคู่ใด ๆ ก็คือ ค่าความเที่ยงของแบบสอบนั้น และอาจกล่าวได้ว่า

ค่าความเที่ยงนี้คือ อัตราส่วนของความแปรปรวนของคะแนนจริง กับความแปรปรวนของคะแนนที่ปรากฏ กล่าวคือ

$$r_{x1x2} = \frac{\sigma_{x^\infty}^2}{\sigma_x^2} = r_{xx}$$

$$\text{หรือ } r_{xx} = \frac{\sigma_{x^\infty}^2}{\sigma_{x^\infty}^2 + \sigma_{xe}^2}$$

ตามแนวคิดของทฤษฎีการสอบเติม กำหนดให้ข้อสอบคู่ขนานมีลักษณะดังนี้

๑. มีค่ามัชฌิมเลขคณิต (\bar{X}) เท่ากัน
๒. มีค่าความเบี่ยงเบนมาตรฐาน (S.D) เท่ากัน
๓. คะแนนของแบบสอบคู่ใด ๆ มีค่าสหสัมพันธ์ (r_{x1x2}) เท่ากัน
๔. คะแนนของแบบสอบใด ๆ มีค่าสหสัมพันธ์ (r_{xy}) กับคะแนนของแบบ

สอบอื่นเท่ากัน

ข. ความเที่ยงตามมโนทัศน์เลือกสรรเกี่ยวกับคะแนนจริงและแบบสอบคู่ขนาน

(Eclectic Concept of True Scores and Parallel Tests)

แนวคิดนี้เชื่อว่า คะแนนจริงของบุคคลหนึ่ง คือ ค่ามัชฌิมเลขคณิตของคะแนนของบุคคลนั้นจากการสอบอนันต์ครั้ง และความแปรปรวนของคะแนนจริงจากการสอบจำนวนหลาย ๆ นี้ มีค่าเท่ากับความแปรปรวนของคะแนนที่ปรากฏ คูณด้วยค่าสหสัมพันธ์ระหว่างแบบสอบคู่ขนาน และค่าสหสัมพันธ์นี้คือ ค่าความเที่ยงของแบบสอบ กล่าวคือ

$$X_{\infty a} = \frac{x_{1a} + \dots + x_{ka}}{k}$$

$$\sigma_{x^\infty}^2 = \sigma_x^2 r_{xx}$$

$$r_{xx} = \frac{\sigma_{x^\infty}^2}{\sigma_x^2}$$

ค. ความเที่ยงตามมโนทัศน์ของการสุ่มข้อสอบจากปริเขต (Concept of Domain Sampling)

แนวคิดนี้เชื่อว่า คะแนนจริงของบุคคลหนึ่ง คือ ค่าเฉลี่ยของคะแนน จากแบบสอบจำนวนมากที่ประกอบด้วยข้อสอบที่สุ่มมาจากปริเขต (domain) ในปริเขต หนึ่งที่มากเพียงพอที่จะเป็นตัวแทนของสิ่งที่วัดได้ ถ้าข้อสอบมีจำนวนเท่ากับข้อสอบทั้งหมด ของปริเขตแล้ว ค่าคะแนนที่ได้คือคะแนนปริเขต และเชื่อว่าค่ามัชฌิมเลขคณิตและค่า ความแปรปรวนของแบบสอบที่ประกอบด้วยข้อสอบที่สุ่มมาจากปริเขตของข้อสอบชุดใด ๆ มีค่าเท่ากัน และค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนของแบบสอบคู่ขนานคู่ใด ๆ ก็คือ ค่าความเที่ยงของแบบสอบนั่นเอง ในทางปฏิบัตินั้น ค่าความเที่ยงอาจหาได้จากวิธีการ ต่าง ๆ ดังนี้ (Ghiselli, ๑๙๖๔:๒๗๗; Stanley, ๑๙๗๑:๔๐๔) คือ

๑. จากการสอบแบบสอบเดิมซ้ำ เช่น วิธีการสอบแล้วสอบซ้ำ (test-retest method) และหาค่าสหสัมพันธ์ของคะแนนการสอบนั้น
๒. จากการสอบแบบสอบฉบับคู่ขนาน (parallel forms) เช่น การสอบแบบสอบชุดหนึ่ง แล้วทำการสอบกับแบบสอบอีกชุดหนึ่ง ที่ประกอบด้วยเนื้อหา อย่างเดียวกัน และมีรูปแบบอย่างเดียวกัน แล้วหาค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนน การสอบนั้น
๓. จากการสอบส่วนหนึ่งของแบบสอบ กับอีกส่วนหนึ่ง เช่น วิธีการ แบ่งครึ่ง (split-half method)
๔. จากความสัมพันธ์ของส่วนต่าง ๆ ของแบบสอบ เช่น Kuder-Richardson Method แบบ KR_{20} KR_{21} Cronbach's α และการวิเคราะห์ความแปรปรวนของผลการสอบตามวิธีของ Hoyt เป็นต้น

จากวิธีการต่าง ๆ ดังกล่าวนี้นี้ การหาค่าความเที่ยงแบบ Cronbach's α และแบบวิธีของ Hoyt สามารถใช้ได้กับการสอบที่คิดคะแนนเป็นแบบ ๐, ๑ หรือคะแนน แบบอื่นก็ได้ และค่าที่คำนวณได้จากทั้งสองวิธีนี้เท่ากัน (Stanley, ๑๙๗๑:๔๑๒)

๒. ความหมายของค่าความเที่ยง

ความเที่ยงของแบบสอบอิงกลุ่ม มีความหมายสำคัญ ๓ ประการ (Ebel, ๑๙๗๒:๔๐๔; Stanley, ๑๙๗๑:๓๗๔) คือ

ก. ความเที่ยง หมายถึง ความคงที่ของผลการสอบของแบบสอบฉบับใดฉบับหนึ่งที่ต้องการวัดสิ่งใดสิ่งหนึ่ง

ข. ความเที่ยง หมายถึง ร้อยละของความแปรปรวนของสิ่ง
ที่วัดที่แบบสอบนั้นสามารถวัดได้

ค. ความเที่ยง หมายถึง ค่าความสัมพันธ์ระหว่างผลทดสอบจากแบบสอบคู่ขนานที่วัดสิ่งเดียวกันคู่ใดคู่หนึ่ง

๓. ปัจจัยสำคัญที่มีผลต่อค่าความเที่ยงของแบบสอบ

ปัจจัยที่สำคัญที่ทำให้ค่าความเที่ยงเปลี่ยนแปลงได้มี ๒ อย่าง (Magnuson, ๑๙๖๖:๕๕) คือ

ก. ความยาวของแบบสอบ ถ้าแบบสอบมีความยาวเพิ่มขึ้น
ค่าความเที่ยงจะเพิ่มขึ้น

ข. การกระจายของคะแนน ถ้าคะแนนของแบบสอบแตกต่างกันมาก ซึ่งอาจเกิดจากข้อสอบยากมาก หรือปัจจัยอื่น ๆ จะทำให้ค่าความเที่ยงเพิ่มขึ้น

ค. ความเที่ยงของแบบสอบอิงเกณฑ์

แนวคิดเกี่ยวกับความเที่ยงของแบบสอบอิงเกณฑ์นั้น มีหลายอย่าง
แต่อาจสรุปได้เป็น ๓ แนวคิด ดังนี้ (Berk, ๑๙๘๐:๓๒๓-๓๔๖; Subkoviak, ๑๙๘๐:
๑๒๗-๑๕๐) คือ

ก. ความเที่ยงตามแนวคิดของ "ฟังก์ชันการสูญเสียน้อยที่สุด"

(Threshold Loss Function)

ความเที่ยงตามแนวคิดนี้ หมายถึง ความคงที่ของการตัดสินใจในการแบ่งผู้สอบออกเป็น ๒ กลุ่ม คือ กลุ่มผู้รู้แล้ว และกลุ่มผู้ยังไม่รู้ ว่าถูกต้องมากน้อยเพียงใด แนวคิดนี้มีสมมุติฐานเบื้องต้น ๒ อย่าง คือ

๑. แบบสอบแบ่งผู้สอบออกเป็น ๒ กลุ่ม คือ กลุ่มผู้รู้แล้ว และกลุ่มผู้ยังไม่รู้ ด้วยเกณฑ์ที่วัดความรู้หรือความสามารถในระดับที่น้อยที่สุดเท่าที่จะสามารถยอมรับได้ หรือการอาศัยจุดตัดที่กำหนดขึ้น

๒. ความสูญเสียหรือความผิดพลาดที่เกิดจากการจำแนกผู้สอบผิดประเภท ไม่ว่าจะกลุ่มใดก็ตามมีความสำคัญเท่าเทียมกัน ไม่ว่าจะความผิดพลาดนั้นจะเท่ากันหรือไม่ก็ตาม

การคำนวณหาค่าความเที่ยงตามแนวคิดนี้มีหลายวิธี เช่น Carver Method, Swaminathan-Hambleton-Algina Method, Huyh Method เป็นต้น บางวิธีอาศัยการสอบ ๒ ครั้ง แต่บางวิธีใช้การสอบเพียง ๑ ครั้ง วิธีต่าง ๆ เหล่านี้โดยมากมีข้อดกลงเบื้องต้นว่า การกระจายของคะแนนของแบบสอบเป็นแบบทรินาม (binomial distribution) และทุกวิธีกำหนดการให้คะแนนด้วยวิธีประเพณีนิยม การหาค่าความเที่ยงโดยวิธีต่าง ๆ นี้ โดยมากจะอาศัยการคำนวณที่ยุ่งยาก และค่าที่คำนวณได้เป็นค่าอคติ (biased) ยกเว้นค่าความเที่ยงที่ได้จาก Swaminathan-Hambleton-Algina Method แต่วิธีนี้ต้องอาศัยการสอบ ๒ ครั้ง (Subkoviak, ๑๙๘๐:๑๕๐)

ข. ความเที่ยงตามแนวคิดของ "ฟังก์ชันการสูญเสียเกิดจากความคลาดเคลื่อนกำลังสอง" (Squared-error Loss Function)

ความเที่ยงตามแนวคิดนี้ หมายถึง ความคงที่ของคะแนนสอบของแต่ละบุคคลว่า แตกต่างจากคะแนนของจุดตัดของแบบสอบคู่ขนานแบบเดิมตามทฤษฎีการสอบเดิม หรือของแบบสอบคู่ขนานแบบสุ่มตามทฤษฎีการอ้างอิงสรุป แนวคิดนี้มีสมมุติฐานเบื้องต้น ๒ อย่าง คือ

๑. คะแนนสอบแต่ละคะแนน คือ ระดับความสามารถของ
ผู้สอบ ซึ่งถือว่ามีความต่อเนื่องกัน

๒. ความสูญเสียหรือความผิดพลาดที่เกิดจากจำแนกผู้สอบ
ให้ห่างจากจุดตัด ถือว่าเป็นความผิดพลาดที่มีความสำคัญมากกว่าการจำแนกให้อยู่ใกล้กับ
จุดตัด

การคำนวณหาความเที่ยงตามแนวคิดนี้มีอยู่ ๒ วิธี คือ
Livingston Method (Livingston, ๑๙๗๒:๒๖) และ Brennan and Kane
Method (Brennan and Kane, ๑๙๗๗:๒๘๓) ทั้ง ๒ วิธีนี้มีความคล้ายคลึงกัน และ
แตกต่างกันดังนี้ (Berk, ๑๙๘๐:๓๓๕) คือ

ก. ความคล้ายคลึงกัน

๑. ไม่กำหนดลักษณะการกระจายของคะแนนการสอบ
๒. ไม่ได้คำนึงถึงความผิดพลาดที่เกิดจากการเดา
๓. ค่าความเที่ยงเป็นอิสระจากค่าความคลาดเคลื่อน
๔. ค่าความเที่ยงเปลี่ยนแปลงไปตามค่าของจุดตัด
๕. เมื่อความยาวของแบบสอบเพิ่มขึ้น ค่าความเที่ยง

ในการวัด

จะเพิ่มขึ้น

ข. ความแตกต่างกัน

๑. Livingston Method อาศัยแนวคิดตามทฤษฎี
การสอบเดิม ในการพัฒนาสูตรการคำนวณ โดยเฉพาะเรื่องแบบสอบคู่ขนานแบบเดิม แต่
Brennan and Kane Method อาศัยแนวคิดตามทฤษฎีการอ้างอิงสรุปในการพัฒนาสูตรการ
คำนวณ และใช้แนวคิดแบบสอบคู่ขนานแบบสุ่ม

๒. Livingston Method ใช้ได้กับการสอบที่
ให้คะแนนตามวิธีประเพณีนิยม แต่ Brennan and Kane Method ใช้ได้กับการสอบที่ให้
คะแนนตามวิธีประเพณีนิยม หรือการให้คะแนนวิธีอื่น ๆ ที่ให้คะแนนข้อสอบแต่ละข้อต่างกัน

๓. ถ้าคะแนนจุดตัด เท่ากับค่ามัชฌิม เลขคณิตของ
การสอบแล้ว ค่าความเที่ยงจาก Livingston Method จะ เท่ากับค่า KR_{20} แต่ค่าความ
เที่ยงจาก Brennan and Kane Method จะ เท่ากับค่า KR_{21}

อนึ่ง สำหรับการศึกษาคครั้งนี้ ผู้วิจัย เลือกใช้การ
หาค่าความเที่ยงแบบ Brennan and Kane Method เพราะข้อตกลงเบื้องต้นสอดคล้อง
กับข้อมูลในการศึกษานี้ และจะได้นำกล่าวถึงวิธีดังกล่าวนี้เพิ่มเติมต่อไปในภายหลัง

ค. ความเที่ยงตามแนวคิดการประมาณคะแนนปริเขต (Domain Score Estimation)

ความเที่ยงตามแนวคิดนี้ หมายถึง ความคงที่ของคะแนนสอบของแต่ละบุคคล
จากการสอบแบบสอบคู่ขนานแบบเดิม หรือแบบสอบคู่ขนานแบบสุ่ม การคำนวณหาความเที่ยง
ตามแนวคิดนี้มีอยู่หลายวิธี เช่น ๑) ใช้ค่าความคลาดเคลื่อนมาตรฐานของการวัด ๒) ใช้
ค่าความคลาดเคลื่อนมาตรฐานของอัตราส่วน ๓) ใช้ค่าเฉลี่ยของความคลาดเคลื่อนมาตรฐาน
ของการวัด และ ๔) ใช้อัตราส่วนระหว่างความแปรปรวนของคะแนนมวลของสิ่งที่วัด กับ
ความแปรปรวนของคะแนนที่สังเกตได้ เป็นต้น

แต่อย่างไรก็ตาม ความเที่ยงตามแนวคิดการประมาณคะแนนปริเขตนี้
ไม่ค่อยมีประโยชน์ต่อการวัดผลในห้องเรียนมากนัก เพราะโดยมากให้ค่าอคติ ค่าเปลี่ยนแปลง
ได้ง่ายตามความยาวของแบบสอบ การคำนวณยุ่งยาก และค่าที่ได้ตีความลำบาก เป็นต้น
แต่ก็มีประโยชน์มากในการประเมินผลโครงการ (Berk, ๑๙๘๐: ๓๔๖)

๒. ความเที่ยงของแบบสอบอิง เกณฑ์ตามแนวคิดของทฤษฎีการอ้างอิง

ดังได้กล่าวมาแล้วว่า การหาค่าความเที่ยงของแบบสอบอิง เกณฑ์นั้น อาจหาได้จากอาศัยแนวคิดของทฤษฎีการอ้างอิงสรุป การศึกษาครั้งนี้ผู้เขียนได้ใช้ค่าดังกล่าวมาใช้เพื่อการศึกษาเปรียบเทียบ เพราะลักษณะข้อมูลสอดคล้องกับแนวคิดดังกล่าว สำหรับทฤษฎีการอ้างอิงสรุปนั้น มีแนวคิดโดยสังเขปดังนี้ (Cronbach and Others, ๑๙๗๒: ๑๐)

ก. แบบสอบในการสอบไม่จำเป็นต้องเป็นแบบสอบคู่ขนานแบบเต็ม คือ ไม่จำเป็นต้องมีค่ามัธยฐาน เลขคณิต และความแปรปรวนเท่ากัน เป็นต้น

ข. ข้อสอบที่ใช้ในการสอบเป็นเพียงตัวอย่างของข้อสอบทั้งหมดที่ได้มาจากการสุ่มแบบง่าย หรือการสุ่มแบบแบ่งชั้น

ค. ทำการพิจารณาผลการสอบ ๒ มิตติ หรือมากกว่าในขณะเดียวกัน

นอกจากนี้ แนวคิดส่วนหนึ่งของทฤษฎีการอ้างอิงสรุป คือ การมุ่งหาความเที่ยงจากการสอบกลุ่มตัวอย่างของสิ่งใดสิ่งหนึ่ง แล้วอ้างอิงสรุปไปยังมวลของสิ่งที่วัด และเรียกดัชนีของการวัดนี้ว่า ค่าสัมประสิทธิ์ของความสามารถในการอ้างอิงสรุป (Coefficient of Generalizability) ซึ่งได้แก่ "อัตราส่วนของความแปรปรวนของคะแนนมวลของสิ่งที่วัด กับความแปรปรวนของคะแนนที่ปรากฏซึ่งคาดหวังไว้" (Cronbach and Others, ๑๙๗๒: ๑๗) และคะแนนมวลของสิ่งที่วัด (Universe Score) คือ ค่าเฉลี่ยของคะแนนจากการวัดภายใต้เงื่อนไขทุกอย่างที่เป็นไปได้ (Brennan, ๑๙๘๓: ๔) คะแนนมีลักษณะเป็นคะแนนในอุดมคติ และมีแนวคิดคล้ายกับแนวคิดของคะแนนจริงในทฤษฎีการสอบเดิม แต่ทว่าคะแนนมวลของสิ่งที่วัดมิได้หลายค่าแล้วแต่เงื่อนไขของการวัด จึงต่างกับคะแนนจริง ซึ่งมีเพียงค่าเดียวเท่านั้น การวัดในตามทฤษฎีการอ้างอิงสรุปนั้น คำนี้ถึงความคลาดเคลื่อนในการวัดซึ่งเกิดจากความแตกต่างระหว่างคะแนนมวลของสิ่งที่วัด กับคะแนนที่ปรากฏซึ่งคาดหวังไว้ ($\Delta \equiv \bar{x}_p - \mu_p$) ซึ่งเรียกว่า ความแปรปรวนของความคลาดเคลื่อนสัมบูรณ์ (absolute error variance: $\sigma^2(\Delta)$) การวัดที่ดีนั้น ค่าดังกล่าวนี้ควรมีน้อยกว่าค่าความแปรปรวนของคะแนนมวลของสิ่งที่วัด

ในปี ๑๙๗๗ Brennan และ Kane (Brennan and Kane, ๑๙๗๗:๒๗๗-๒๘๓) ได้เสนอแนวคิดเกี่ยวกับการหาค่าความเที่ยงของแบบสอบถามเชิงทัศนคติ โดยอาศัยแนวคิดจากทฤษฎีการอ้างอิงสรุป และเรียกค่าความเที่ยงนี้ว่า ดัชนีความเชื่อถือ (index of dependability) หรือบางทีเรียกว่า ค่าสัมประสิทธิ์ที่มีลักษณะคล้ายกับค่าความเที่ยง (reliability-like coefficient) เพราะว่า แนวคิดของค่าดังกล่าวนี้คล้ายกับแนวคิดของค่าความเที่ยงของแบบสอบถามกลุ่ม แต่พัฒนาการของสูตรอาศัยทฤษฎีการสอบต่างกัน ซึ่งจะสังเกตได้จากสูตรในการคำนวณต่อไปนี้ (Brennan, ๑๙๘๑:๖๘)

$$\Phi(\lambda_0) = \frac{\sigma^2(\pi) + (\mu - \lambda_0)^2}{\sigma^2(\pi) + (\mu - \lambda_0)^2 + \sigma^2(\Delta)}$$

จากสมการข้างบนนี้จะเห็นได้ว่า ค่าดัชนีความเชื่อถือ คือ "อัตราส่วนของความแปรปรวนของคะแนนรวมของสิ่งที่วัด [$\sigma^2(\pi)$] และค่ากำลังสองของความแตกต่างของ $(\mu - \lambda_0)$ กับค่าดังกล่าวแล้วทั้งหมด และค่าความแปรปรวนของความคลาดเคลื่อนในการวัด [$\sigma^2(\Delta)$]" จะเห็นได้ว่า ค่าที่เป็นเศษนั้นคล้ายกับค่า σ_∞^2 และค่าที่เป็นส่วนนั้นคล้ายกับค่า $\sigma_\infty^2 + \sigma_e^2$ ตามแนวคิดของทฤษฎีการสอบเดิม ซึ่งอัตราส่วนของค่าดังกล่าวนี้ก็คือ ค่าความเที่ยง (Ghiselli, ๑๙๖๔:๒๒๕-๖) ค่า $\Phi(\lambda_0)$ มีค่าน้อยที่สุดเมื่อจุดตัด (λ_0) เท่ากับค่าเฉลี่ยของการสอบ และมีค่าเท่ากับค่า KR_{21} แต่เมื่อความแตกต่างของค่าทั้งสองมากขึ้นหรือน้อยลง ค่า $\Phi(\lambda_0)$ จะเพิ่มขึ้น ส่วนค่า $\sigma^2(\Delta)$ จะคงเดิม

Brennan (Brennan, ๑๙๗๘:๓๘; Brennan and Kane, ๑๙๗๗:๒๘๑) อธิบายว่า การที่ค่า $\Phi(\lambda_0)$ มีเพิ่มขึ้นเมื่อคะแนนจุดตัดต่างจากคะแนนค่าเฉลี่ย แสดงว่า ผู้สอบมีคะแนนต่างจากค่าคะแนนเฉลี่ยยิ่งมากก็ยังสามารถถูกแบ่งให้อยู่ในกลุ่มผู้รู้แล้ว หรือกลุ่มผู้ยังไม่รู้ ได้ถูกต้องยิ่งขึ้น ส่วนการที่ค่า $\sigma^2(\Delta)$ คงเดิมนั้น เป็นเพราะว่าคะแนนของจุดตัดมีผลต่อค่า $\Phi(\lambda_0)$ ในส่วนที่เป็นเศษ ซึ่งถือว่าเป็นสาระสำคัญมากกว่าส่วนที่เป็นความคลาดเคลื่อน

การคำนวณหาค่าความเที่ยงตามแนวคิดของทฤษฎีอ้างอิงสรุปนี้ ถ้าอาศัยค่า $\sigma^2(\pi)$ และค่า $\sigma^2(\delta)$ เท่านั้น จะได้ค่าเท่ากับค่า KR₂₀ หรือค่า Cronbach's α (Brennan and Kane, ๑๙๗๗:๒๘๐; Brennan, ๑๙๘๐:๒๐๒) และค่าดังกล่าวนี้เรียกว่า ค่าสัมประสิทธิ์ของความสามารถในการอ้างอิงสรุป (Generalizability Coefficient) คำนวณได้ดังนี้

$$\Sigma\rho^2 = \frac{\sigma^2(\pi)}{\sigma^2(\pi) + \sigma^2(\delta)}$$

อนึ่ง ค่า $\Phi(\lambda_0)$ มีความหมายดังนี้ (Brennan, ๑๙๗๙:๓๔)

๑. เป็นดัชนีความคงที่ของคะแนนสอบของผู้สอบแต่ละบุคคลว่า แตกต่างจากจุดตัด ซึ่งใช้ เป็นเกณฑ์ในการสอบว่ามีมากน้อยเพียงใด
๒. เป็นค่าดัชนีของความสอดคล้องระหว่างคะแนนของแบบสอบคู่ขนานแบบ สุ่ม ๒ ฉบับ ของการสอบถึง เกณฑ์ของผู้สอบแต่ละคนว่ามีมากน้อยเพียงใด

การคำนวณค่า $\Phi(\lambda_0)$ มีข้อตกลงเบื้องต้นดังนี้ (Brennan, ๑๙๘๐:๒๒๓; Berk, ๑๙๘๐:๓๓๖; Brennan and Kane, ๑๙๗๗:๒๗๗)

๑. ข้อสอบมาจากขอบข่ายของเนื้อหาที่นิยามไว้อย่างดีแล้ว
๒. ข้อสอบในการวัดได้จากการสุ่มอย่างง่าย หรือการสุ่มแบ่งชั้นจากมวลของสิ่งที่วัดจำนวนมาก แต่นำตัวอย่างมาศึกษาเพียงจำนวนน้อย
๓. แบบสอบคู่ขนานเป็นแบบสอบคู่ขนานแบบสุ่ม ไม่ใช่แบบสอบคู่ขนานแบบเดิม
๔. ให้คะแนนการสอบโดยวิธีประเพณีนิยม หรือวิธีให้คะแนนข้อสอบต่างกัน
๕. การกระจายของคะแนนไม่จำกัดรูปแบบ

อนึ่ง จากแนวคิดทั้ง ๓ ข้อของทฤษฎีการอ้างอิงสรุปดังกล่าวแล้วข้างต้น ทำให้วิธีการคำนวณหาค่าความสามารถในการอ้างอิงสรุป ต้องอาศัยแนวคิดของการวิเคราะห์ความแปรปรวนในรูปแบบต่างกัน แล้วแต่ลักษณะเงื่อนไขการสอบ เช่น อาจเป็นรูปแบบการสุ่ม

(random model) หรือรูปแบบผสม (mixed model) เป็นต้น (Brennan, ๑๙๘๓:๔) แต่ว่าวิธีการคำนวณต่างกับ F-test เพราะมุ่งที่ส่วนประกอบต่าง ๆ ของความแปรปรวน ที่ได้จากการประมาณมากกว่าความแปรปรวนต่าง ๆ ที่ได้จากการวัดมวลของสิ่งที่วัด ด้วยเหตุนี้ การวิเคราะห์ความแปรปรวนจึงต้องเกี่ยวกับการศึกษาความสามารถในการอ้างอิงสรุป (Generalizability study: G study) และการศึกษาการตัดสินใจ (Decision study: D study) เพิ่มขึ้น การศึกษาทั้ง ๒ อย่างนี้ มีแนวคิดโดยสังเขป ดังนี้

๑. การศึกษาความสามารถในการอ้างอิงสรุป (G study)

ในขั้นนี้เป็นการศึกษามวลของข้อมูลที่สามารถหาได้ (universe of admissible observations) โดยการวิเคราะห์ความแปรปรวนจากแหล่งต่าง ๆ ต่อจากการวิเคราะห์ความแปรปรวนโดยวิธีปกติทั่วไป ค่าที่ได้จากการคำนวณเป็นค่าประมาณของค่าของมวลสิ่งที่วัด หรือค่าของประชากรของบุคคลที่รับการสอบ เช่น $\hat{\sigma}^2(p)$ หรือ $\hat{\sigma}^2(i)$ เป็นต้น การคำนวณหาค่าต่าง ๆ เหล่านี้ เป็นจุดประสงค์ที่สำคัญของการศึกษาในขั้นนี้ และค่าต่าง ๆ ในขั้นนี้เป็นค่าอคติ

๒. การศึกษาการตัดสินใจ (D study)

จุดมุ่งหมายในการศึกษาขั้นนี้ เพื่อการประมาณค่าความแปรปรวนต่าง ๆ ที่จะใช้เพื่อการตัดสินใจในการอ้างอิงสรุปค่าที่ได้ไปยังค่ามวลของสิ่งที่วัด หรือค่าประชากร ดังนั้น ค่าที่ได้ในขั้นนี้จึงเป็นค่าพารามิเตอร์ ซึ่งมีลักษณะของค่าในอุดมคติดังได้กล่าวแล้ว

๓. ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory)

ทฤษฎีการตอบสนองข้อสอบนี้ บางทีเรียกว่า ทฤษฎีความสามารถแฝง (Latent Trait Theory) หรือทฤษฎีโค้งคุณลักษณะของข้อสอบ (Item Characteristic Curve Theory) เป็นทฤษฎีการวัดคุณลักษณะภายในข้อสอบ โดยการวิเคราะห์ข้อสอบรายข้อ เพื่อใช้ประกอบการศึกษา ทฤษฎีนี้เชื่อว่าผลการสอบของผู้สอบจากแบบสอบใด ๆ ขึ้นอยู่กับความสามารถ หรือทักษะของผู้สอบ กับคุณลักษณะของข้อสอบ เช่น ค่าความยากง่าย

ค่าอำนาจจำแนก และค่าการเดา ความสามารถหรือทักษะของคน กับค่าต่าง ๆ เหล่านี้ มีความสัมพันธ์กันในเชิงคณิตศาสตร์ในรูปของฟังก์ชัน ซึ่งจะมีค่าพารามิเตอร์เกี่ยวข้องกับตัวที่ตัว ขึ้นอยู่กับพฤติกรรมการตอบสนองข้อสอบของผู้สอบ (Warm, ๑๙๗๙:๑๙) Lord and Novick, ๑๙๖๘:๓๖๘)

ปกติแล้ว ฟังก์ชันของทฤษฎีการตอบสนองข้อสอบมีหลายอย่าง แต่ที่นิยมใช้กันมาก มี ๓ รูปแบบ (Hambleton and Cook, ๑๙๗๗:๘๐-๘๒) คือ

๑. รูปแบบโลจิสติกใช้พารามิเตอร์ ๓ ตัว (Three-Parameter Logistic Model) ใช้ค่าอำนาจจำแนก (a) ค่าความยากง่าย (b) และค่าการเดา (c) มีสูตรการคำนวณดังนี้

$$p_g(\theta) = c_g + (1-c_g) + \frac{\text{EXP}[D_{a_g}(\theta-b_g)]}{1+\text{EXP}[D_{a_g}(\theta-b_g)]}$$

ในเมื่อ $D = \text{scaling factor}$

$\text{EXP} = ๒.๗๑๘๒๘๑๘$

$\theta = \text{ความสามารถของผู้สอบ}$

๒. รูปแบบโลจิสติกใช้พารามิเตอร์ ๒ ตัว (Two-Parameter Logistic Model) ใช้ค่าอำนาจจำแนก (a) และค่าความยากง่าย (b) เท่านั้น มีสูตรการคำนวณดังนี้

$$p_g(\theta) = \frac{\text{EXP}[D_{a_g}(\theta-b_g)]}{1+\text{EXP}[D_{a_g}(\theta-b_g)]}$$

๓. รูปแบบโลจิสติกใช้พารามิเตอร์ ๑ ตัว (One-Parameter Logistic Model) หรือรูปแบบราสช์ (Rasch Model) เพราะ Georg Rasch เป็นผู้เสนอขึ้นใช้ในปี ๑๙๖๐ (Rasch, ๑๙๘๐) วิธีนี้ใช้ค่าความยากง่ายเพียงค่าเดียว มีสูตรการคำนวณดังนี้ [สำหรับรูปแบบราสช์บางตำราใช้ β แทน b งานวิจัยนี้ใช้ β]

$$p_g(\theta) = \frac{\text{EXP}(\theta - \beta_g)}{1 + \text{EXP}(\theta - \beta_g)}$$

ตามทฤษฎีแล้ว ค่า θ a และ b ของแต่ละวิธีมีค่าระหว่าง $-\infty$ ถึง $+\infty$ แต่ในทางปฏิบัติ ค่า θ มีค่าระหว่าง -3 ถึง $+3$ ค่า a มีค่าระหว่าง 0.5 ถึง 2.5 และค่า b มีค่าระหว่าง -2.5 ถึง 2.5 ส่วนค่า c นั้น มีค่าระหว่าง 0 ถึง 0.3 เมื่อข้อสอบมี ๔ ตัวเลือก (Ree, ๑๙๗๔:๓๗๒; Warm, ๑๙๗๘:๕๒-๕๓)

อนึ่ง ข้อตกลงเบื้องต้นที่สำคัญของทฤษฎีการตอบสนองข้อสอบมีดังนี้ (Lord and Novick, ๑๙๖๘:๓๕๔; Hambleton and Cook, ๑๙๗๗:๗๘)

๑. แบบสอบวัดมิติเดียว (unidimension) หมายความว่า ข้อสอบแต่ละข้อของแบบสอบนั้น วัดความสามารถหรือคุณลักษณะ หรือทักษะอย่างใดอย่างหนึ่งเพียงอย่างเดียวเท่านั้น หรือข้อสอบเหล่านี้มีความเป็นเอกพันธ์

๒. ข้อสอบแต่ละข้อเป็นอิสระจากกัน (local independence) หมายความว่า การตอบข้อสอบข้อใดข้อหนึ่งถูกหรือไม่ ไม่ขึ้นอยู่กับคำตอบข้อสอบข้ออื่น ๆ ข้อสอบจะอยู่ในตำแหน่งใดของแบบสอบก็ได้ จะไม่มีผลต่อการตอบ

๓. โอกาสที่ผู้ตอบจะตอบข้อสอบถูกหรือไม่ ขึ้นอยู่กับลักษณะของโค้งการตอบข้อสอบ (item characteristic curve) ของแต่ละรูปแบบที่ใช้ ไม่ใช่ขึ้นอยู่กับลักษณะการแจกแจงความสามารถของกลุ่มตัวอย่าง

สำหรับการใช้แต่ละรูปแบบ ยังมีข้อตกลงเบื้องต้นแตกต่างกันบ้าง เพราะบางวิธีก็มีข้อจำกัดบางอย่าง และบางวิธีก็มีความแกร่ง (Robustness) เฉพาะ เช่น การวัดให้ได้ผลการคำนวณจากรูปแบบโลจิสติกส์ใช้พารามิเตอร์ ๓ ตัว ควรใช้กับกลุ่มตัวอย่างขนาดใหญ่ เช่น ประมาณ ๑,๐๐๐ คน และข้อสอบควรมีจำนวนมาก เช่น ประมาณ ๔๐ ข้อ (Wood and Others, ๑๙๗๖:๕) แต่การใช้รูปแบบราสนั้น กลุ่มตัวอย่างประชากรมีขนาดเล็กประมาณ ๑๐๐ คน ก็ใช้ได้ แต่ถ้าจะให้ผลดีควรมีประมาณ ๔๐๐ คน และข้อสอบนั้นยาวเพียง ๒๐ ข้อ ก็ใช้ได้ แต่ถ้าจะให้ผลดีควรมีประมาณ ๔๐ ข้อ (Wright, ๑๙๗๗: ๒๑๙; Wright and Mead, ๑๙๗๘:๒๓-๒๕) และอำนาจจำแนกของข้อสอบนั้นไม่จำเป็นต้องเท่ากันก็ได้ (Wright and Panchapakesan, ๑๙๖๘:๒๕)

อนึ่ง ค่าพารามิเตอร์ต่าง ๆ ที่ได้จากการวิเคราะห์ข้อสอบรายข้อจากรูปแบบต่าง ๆ ดังกล่าวแล้ว เป็นค่าที่ไม่เปลี่ยนแปลงไปตามกลุ่มตัวอย่าง และไม่เปลี่ยนแปลงไปตามตำแหน่งของข้อสอบ คุณลักษณะดังกล่าวนี้คือ ความเป็นปรนัยของการวัดผล (objectivity of measurement) ที่มีความสำคัญมากในทางวัดผล (Wright and Stone, ๑๙๗๘:๑๕)

จ. ความแตกต่างระหว่างแบบสอบอิงกลุ่มและแบบสอบอิง เกณฑ์

แบบสอบอิง เกณฑ์เป็นแบบสอบที่นักทดสอบ เพิ่งให้ความสนใจอย่างจริงจังเมื่อประมาณปี ๑๙๗๐ มา นี้ โดย Glaser และ Nitko ได้เป็นผู้เสนอแนวคิดไว้ว่า "แบบสอบอิง เกณฑ์เป็นแบบสอบที่สร้างขึ้นอย่างพิถีพิถันเพื่อการวัดผล ซึ่งสามารถที่จะแปลความหมายได้โดยตรงตามมาตรฐานของความสามารถจลน์ (performance) ที่กำหนดไว้" (Glaser and Nitko, ๑๙๗๑:๖๕๓) และอีกคำนิยามหนึ่งที่นักทดสอบมักอ้างอิงเสมอคือ "แบบทดสอบอิง เกณฑ์ คือ แบบทดสอบซึ่งใช้เพื่อวัดสภาพที่แน่นอนของแต่ละบุคคลโดยอาศัยเกณฑ์บางอย่าง เช่น มาตรฐานของความสามารถจลน์ หรือปริเขต (domain) ของพฤติกรรมที่ได้นิยามไว้ อย่างดีแล้ว" (อ้างอิงจาก Hambleton and Others, ๑๙๗๘:๑)

จากคำนิยามทั้งสองดังกล่าวแล้วจะเห็นได้ว่า แบบสอบอิง เกณฑ์มีลักษณะพิเศษ ๒ ประการคือ ๑) ส่วนประกอบของแบบสอบ ซึ่งประกอบด้วยข้อสอบที่สร้างจากวัตถุประสงค์เชิงพฤติกรรม และ ๒) เกณฑ์ของพฤติกรรมหรือเกณฑ์ของความสามารถจนที่ได้กำหนดไว้ก่อนแล้ว ซึ่งตามคำนิยามทั้งสอง หมายถึงกลุ่มของพฤติกรรม ซึ่งอาจจะมีหลายอย่าง หรือเพียงอย่างเดียวก็ได้ แต่ได้กำหนดขอบเขตเนื้อหาไว้อย่างชัดเจนแล้ว (Millman, ๑๙๗๔: ๓๔๕) ดังนั้น แบบสอบอิง เกณฑ์อาจไม่จำเป็นต้องมีจุดตัดของคะแนนก็ได้

อนึ่ง Pilliner (Pilliner, ๑๙๗๔: ๔๔) กล่าวว่า แบบสอบอิง เกณฑ์และแบบสอบอิงกลุ่มมีความแตกต่างกัน ๕ อย่าง คือ

๑. จุดมุ่งหมาย

แบบสอบอิงกลุ่มใช้เพื่อวัดระดับความสามารถของผู้สอบโดยการเปรียบเทียบกับระดับความสามารถของผู้อื่นเป็น เกณฑ์ แต่แบบสอบอิง เกณฑ์ใช้เพื่อวัดระดับความสามารถของผู้สอบโดยการเปรียบเทียบกับระดับความสามารถที่กำหนดไว้ เป็น เกณฑ์

๒. การสร้าง

แบบสอบอิงกลุ่มเน้นที่การเลือกข้อสอบที่มีค่าอำนาจจำแนกสูงไว้ และตัดข้อสอบที่ไม่ดีทิ้ง แต่แบบสอบอิง เกณฑ์จะรวมข้อสอบทุกข้อที่วัดสิ่งที่ต้องการวัดไว้

๓. การใช้

แบบสอบอิงกลุ่มใช้เพื่อการคัดเลือก และเพื่อการเปรียบเทียบระดับความรู้ความสามารถ แต่แบบสอบอิง เกณฑ์ใช้เพื่อการวินิจฉัยความรู้ความสามารถของผู้สอบ

๔. การสารสนเทศที่ได้

แบบสอบอิงกลุ่มให้สารสนเทศเกี่ยวกับผู้สอบกว้าง ๆ แต่แบบสอบอิง เกณฑ์ให้สารสนเทศเกี่ยวกับผู้สอบได้อย่าง เฉพาะ

๕. การอ้างอิงสรุป

ผลการสอบของแบบสอบอิงกลุ่มสามารถใช้อ้างอิงสรุปเกี่ยวกับความรู้ความสามารถของผู้สอบได้อย่างจำกัด และต้องใช้อ้างอิงกับค่าความตรงของแบบสอบ แต่ผลการสอบของแบบสอบอิง เกณฑ์สามารถใช้อ้างอิงสรุปได้กว้างกว่า สำหรับเนื้อหาที่เฉพาะเจาะจงกว่า

แต่อย่างไรก็ตาม นักทดสอบยังมีความสับสนว่า แบบสอบอิงเกณฑ์คืออะไร และมีคุณลักษณะอย่างไรบ้าง ดังจะเห็นได้จากการศึกษาของ Hambleton และคณะ ในปี ๑๙๗๘ (Hambleton and Others, ๑๙๗๘:๒) พบว่ามีคำนิยามเกี่ยวกับ "แบบสอบอิงเกณฑ์" ประมาณ ๖๐๐ คำนิยามที่ปรากฏในวรรณคดีต่าง ๆ ทางการศึกษา แต่คำนิยามเหล่านี้แตกต่างกันเท่า ๆ กับจำนวนผู้ที่ให้คำนิยาม นักทดสอบบางท่าน เช่น Millman Hambleton และคณะ (อ้างจาก Block and Dockrell, ๑๙๘๔; Hambleton and Others, ๑๙๗๘:๓) กล่าวว่า "เมื่อจุดมุ่งหมายของสิ่งที่วัดของแบบสอบอิงกลุ่มสอดคล้องกับจุดมุ่งหมายของสิ่งที่วัดของแบบสอบอิง เกณฑ์ แบบสอบอิงกลุ่มสามารถใช้เป็นแบบสอบอิง เกณฑ์ได้และแบบสอบอิง เกณฑ์ก็สามารถใช้เป็นแบบสอบอิงกลุ่มได้" และ "แบบทดสอบโดยมากสามารถใช้เป็นแบบสอบอิงกลุ่มและแบบสอบอิง เกณฑ์ได้ด้วย"

ดังนั้น สิ่งที่แตกต่างกันอย่างเห็นได้ชัดระหว่างแบบสอบอิงกลุ่มและแบบสอบอิง เกณฑ์ ก็คือ "การแปลความหมายของคะแนน" (interpretation) ส่วนลักษณะของเครื่องมือ (nature of the instruments) ของแบบสอบทั้ง ๒ แบบ ยังไม่แตกต่างกันอย่างชัดเจน (Block and Dockrell, ๑๙๘๔:๓๖) ด้วยเหตุนี้ นักทดสอบบางท่านในปัจจุบัน เช่น Brennan (Brennan, ๑๙๘๔:๓๒๙) ผู้ซึ่งมีชื่อเสียงมากทางทฤษฎีการอ้างอิงสรุปเชื่อว่า ความแตกต่างของแบบสอบอิงกลุ่ม และแบบสอบอิง เกณฑ์อยู่ที่การแปลความหมายของคะแนนเท่านั้น ส่วนประกอบของแบบสอบและการคัดเลือกข้อทดสอบนั้นเป็นเพียงสิ่งที่ทำให้แบบสอบสามารถใช้อ้างอิงสรุปผลการสอบได้มากน้อยเพียงใด เท่านั้น ซึ่งแบบสอบทั้ง ๒ แบบ อาจมีความแตกต่างกันหรือเหมือนกันก็ได้

ตอนที่ ๓ งานวิจัยที่เกี่ยวกับปัจจัยที่มีผลต่อความตรงและความเที่ยง

ก. ปัจจัยที่มีผลต่อความตรง

ปัจจัยที่มีผลต่อความตรงของแบบสอบถามมาก ปัจจัยบางอย่างนอกจากมีผลต่อค่าความเที่ยงแล้ว ยังมีผลต่อความตรงของแบบสอบถาม เช่น การเดา และคำสั่งในการทำแบบสอบถาม เป็นต้น ปัจจัยที่สำคัญต่อความตรงมีดังนี้ (Gronlund, ๑๙๗๖:๙๘-๑๐๒)

๑. ปัจจัยที่เกิดจากผู้สอบ

ก. การเดา

จากการศึกษาของ Lord (Lord, ๑๙๖๔:๗๔๔) เมื่อใช้กลุ่มตัวอย่าง ๒,๐๐๐ คน ทำแบบสอบศัพท์ยาว ๒๕ ๕๐ และ ๖๐ ข้อ โดยมีคำสั่งให้ผู้สอบเดาคำตอบได้ถ้าไม่รู้คำตอบ ผลปรากฏว่า การบังคับให้เดาอย่างไม่มีระบบ ทำให้แบบสอบถามมีค่าความตรงเพิ่มขึ้น ถ้าการเดามีมากและผลของการเดาสัมพันธ์กับเกณฑ์ และการบังคับให้เดาอย่างมีระบบ ทำให้แบบสอบถามมีค่าความตรงเพิ่มขึ้นอย่างมีนัยสำคัญด้วยเช่นกัน ดังนั้น Lord จึงแนะนำว่า การใช้คำสั่งให้ผู้สอบเดาคำตอบข้อที่ไม่รู้นั้นควรระวัง เพราะอาจทำให้ค่าความตรงเพิ่มขึ้นได้ แต่จากการศึกษาของ Sax และ Collet (Sax and Collet, ๑๙๖๘:๑๑๒๗) พบว่า แบบสอบที่ห้ามผู้สอบเดาคำตอบ ถ้าเดาผิดจะหักคะแนน และแบบสอบที่ห้ามผู้สอบเดาคำตอบ แต่ถ้าบอกว่าตัวเลือกใดผิดได้ถูกต้อง จะให้คะแนนเพิ่มต่างก็มีค่าความตรงไม่แตกต่างกันอย่างมีนัยสำคัญ จากแบบสอบที่อนุญาตให้ผู้สอบเดาคำตอบได้

ข. กิจนิสัย (response set) ในการทำแบบสอบ

ผู้สอบมักมีนิสัยในการทำแบบสอบต่างกัน เช่น บางคนทำข้อสอบเลือกตอบไม่ได้ จะเลือกข้อ ค เป็นคำตอบที่ถูก บางคนจะเลือกข้อ ก เป็นข้อถูกเสมอ เป็นต้น ผลของการทำแบบสอบเช่นนี้ทำให้ไม่อาจทราบได้ว่า แบบสอบสามารถวัดสิ่งที่ต้องการจะวัดได้มากน้อยเพียงใด

ค. ความไม่พร้อมทางร่างกายและจิตใจ

ความวิตกกังวล ความเจ็บป่วย ความเบื่อหน่าย และสภาพของร่างกายและจิตใจไม่ปกติต่าง ๆ ในระหว่างการสอบ ทำให้การสอบได้ไม่ดีเท่ากับการสอบในสภาวะปกติ

๒. ปัจจัยที่เกิดจากการบริหารการสอบ

ก. การให้คะแนน

การให้คะแนนที่ไม่เป็นปรนัย ทำให้ผลการสอบได้รับการตัดสินไม่แน่นอน จึงทำให้แบบสอบมีความตรงไม่แน่นอน

ข. การบริหารการสอบ

สภาพต่าง ๆ ในการทำแบบสอบมีผลต่อความตรงของการสอบ เช่น ห้องสอบร้อนเกินไป แสงสว่างไม่เพียงพอ มีเสียงดังมาก ผู้บริหารการสอบให้คำแนะนำในการสอบได้ไม่ดี เช่น เสียงไม่ดังพอ หรือทำความเข้าใจกับผู้สอบได้ไม่ชัดเจนเพียงพอ เป็นต้น

ค. ระยะเวลาในการสอบ

ระยะเวลาในการสอบ นอกจากจะมีผลต่อความเที่ยงของแบบสอบแล้ว ยังมีผลต่อความตรงด้วย Boxter (อ้างจาก Kandall, ๑๙๖๔:๗๘๙) พบว่า การให้เวลาในการทำแบบสอบนานเกินไป ทำให้แบบสอบมีความตรงลดลงอย่างมีนัยสำคัญ เขากล่าวว่า เวลาในการสอบนานเกินไป หรือน้อยเกินไป ทำให้ผู้สอบเดาคำตอบข้อที่ทำไม่ได้ ซึ่งมีผลต่อค่าความเที่ยงและความตรงด้วย แต่จากการศึกษาของ Kandall (Kandall, ๑๙๖๔:๗๘๙) โดยใช้กลุ่มตัวอย่าง ๗๓๔ คน ซึ่งเป็นทหารในกองทัพพบพบว่า แบบสอบที่ให้เวลาสอบมากพอดี จะมีค่าความตรงสูงที่สุด แต่ถ้าให้เวลามากเกินไป ค่าความเที่ยงจะคงที่ และถ้าให้เวลาน้อยเกินไป ค่าความเที่ยงจะลดลงอย่างมีนัยสำคัญ

๓. ปัจจัยที่เกิดจากแบบสอบ

ก. คำสั่งในการทำแบบสอบไม่ชัดเจน

คำสั่งที่ไม่ชัดเจนทำให้ผู้สอบเข้าใจผิดได้ง่าย ซึ่งอาจเกิดจากไม่มีตัวอย่างประกอบ ใช้ภาษาไม่รัดกุมและพิมพ์ไม่ชัดเจน เป็นต้น

ข. คำศัพท์และโครงสร้างของภาษายากเกินไป

คำศัพท์และโครงสร้างของภาษาที่ใช้ยากเกินไป อาจทำให้ผู้สอบไม่เข้าใจส่วนที่ไม่เกี่ยวกับสิ่งที่ต้องการสอบ แต่เสียเวลามาก จนทำแบบสอบได้ไม่เสร็จทันเวลา

ค. ข้อสอบมีความยากไม่เหมาะสม

ในกรณีแบบสอบอิงกลุ่ม ข้อสอบที่ยากเกินไปหรือง่ายเกินไป ทำให้แบบสอบมีค่าความตรงต่ำ และในกรณีของการสอบอิงเกณฑ์ ข้อสอบที่ง่ายมากเกินไป อาจไม่เหมาะสมกับจุดมุ่งหมายในการสอบ ดังนั้น จะทำให้แบบสอบมีความตรงต่ำด้วยเช่นเดียวกัน

ง. ข้อสอบไม่ดี

ข้อสอบแบบเลือกตอบ ถ้าสร้างไม่ดีตัวเลือกอาจแนะนำคำตอบได้ โดยผู้ออกข้อสอบไม่ได้ตั้งใจ ทำให้ข้อสอบวัดสิ่งที่ต้องการวัดอย่างแท้จริงไม่ได้

จ. ข้อสอบกำกวม

ข้อสอบที่ใช้ภาษากำกวม อาจทำให้ผู้สอบที่เก่งทำไม่ได้ แต่ผู้สอบที่ไม่เก่งทำได้ และอาจทำให้ผู้สอบเข้าใจความหมายของภาษาผิดพลาดได้

ฉ. ชนิดของแบบสอบที่ใช้ไม่เหมาะสม

แบบสอบที่ไม่เหมาะสม อาจวัดสิ่งที่ต้องการวัดได้ไม่ดีเพียงพอ ทำให้ไม่อาจประเมินสิ่งที่ต้องการวัดได้อย่างถูกต้อง เช่น การสอบความสามารถในการเขียน ก็ควรใช้แบบสอบที่ให้ผู้สอบเขียนมากกว่าจะให้เลือกคำตอบจากแบบสอบแบบเลือกตอบเท่านั้น เป็นต้น

ข. แบบสอบสั้นเกินไป

แบบสอบที่สั้นเกินไปอาจทำให้วัดสิ่งที่ต้องการวัดได้ไม่มาก และไม่ครอบคลุมเนื้อหาที่ต้องวัด ทำให้การประเมินผลการสอบผิดพลาดได้ เพราะสิ่งที่วัดไม่เป็นตัวแทนของเนื้อหาที่จะวัด

นอกจากนี้เป็นที่ทราบกันแล้วว่า ความตรงของแบบสอบขึ้นอยู่กับความเที่ยงและความเที่ยงของแบบสอบขึ้นอยู่กับความยาวของแบบสอบ ซึ่งเป็นปัจจัยที่สำคัญอย่างหนึ่ง ดังนั้น ความตรงของแบบสอบก็ขึ้นอยู่กับความเที่ยงด้วย Guilford ได้เสนอสูตรในการคำนวณค่าความตรงของแบบสอบที่มีความยาวเพิ่มขึ้น n เท่า ดังนี้ (Guilford, ๑๙๕๔:๔๐๗)

$$r_{nx.y} = \frac{r_{xy}}{\sqrt{\frac{1-r_{xx}}{n} + r_{xx}}}$$

ในเมื่อ $r_{nx.y}$ = ค่าความตรงของแบบสอบที่มีความยาวเพิ่มขึ้น
 r_{xy} = ค่าสัมประสิทธิ์สหสัมพันธ์ของแบบสอบกับ เกณฑ์
 n = จำนวนเท่าของข้อสอบที่เพิ่มขึ้น
 r_{xx} = ค่าความเที่ยงของแบบสอบ

ข. กระสวนของคำตอบของแบบสอบ

ถ้าผู้สอบสามารถรู้ได้ว่า กระสวนของคำตอบของแบบสอบเป็นอย่างไร เช่น คำตอบที่ถูกต้อง คือ ข้อ ค ทุกข้อ จะทำให้ผู้สอบเลือกคำตอบตามกระสวนของคำตอบที่ตนคิดว่าถูกต้องแล้ว ทำให้การวัดไม่อาจวัดสิ่งที่ต้องการจะวัดได้อย่างแท้จริง ดังนั้นแบบสอบจะมีความตรงต่ำ

๔. ปัจจัยเกิดจากการสอนของครู

ปัจจัยที่เกิดจากครูที่มีผลต่อความตรงของแบบสอบก็คือ เนื้อหาที่สอน กับสิ่งที่สอบไม่สัมพันธ์กัน เช่น ถ้าต้องการสอบความรู้ความสามารถในการคำนวณ แต่คำตอบ และวิธีการได้เฉลยให้ผู้สอบรู้แล้ว จึงวัดได้แต่ความจำเท่านั้น หรือสอบสิ่งที่ไม่ได้สอนเลย เป็นต้น

๕. ปัจจัยที่เกิดจากกลุ่มตัวอย่างและเกณฑ์ในการอ้างอิง

ก. กลุ่มตัวอย่าง

ค่าความตรงและค่าความเที่ยงเป็นค่าของแบบสอบที่เกิดจากการสอบแต่ละครั้ง จะเปลี่ยนแปลงไปตามลักษณะของผู้สอบ เช่น ถ้าผู้สอบมีจำนวนมาก ความรู้ความสามารถแตกต่างกัน ค่าความตรงหรือความเที่ยงก็จะมีค่าอย่างหนึ่ง แต่ถ้ากลุ่มตัวอย่างมีขนาดเล็ก ความสามารถไม่แตกต่างกันมาก ก็จะทำให้แบบสอบมีค่าความตรงและความเที่ยงเปลี่ยนไป นอกจากนี้ เพศ อายุ พื้นฐานทางการศึกษา และพื้นฐานทาง เศรษฐกิจและสังคม ก็มีผลทำให้แบบสอบมีค่าความตรงและความเที่ยงแตกต่างกันได้

ข. เกณฑ์ในการอ้างอิง

ค่าความตรงของแบบสอบที่ต้องอาศัยเกณฑ์ในการอ้างอิง จะมีค่าแตกต่างกัน แล้วแต่ความสัมพันธ์ระหว่างเกณฑ์กับแบบสอบนั้นว่า จะมีมากน้อยเพียงใด ในการเลือกเกณฑ์นั้น ควรพิจารณาว่าตามความเป็นจริงแล้ว เกณฑ์กับแบบสอบนั้นสามารถวัดสิ่งที่ต้องการวัดได้มากน้อยเพียงใด เช่น แบบสอบวัดความสามารถในการคิดเชิงวิพากษ์วิจารณ์ (critical thinking) สัมพันธ์กับผลการเรียนวิชาทางสังคมศาสตร์มากกว่าวิชาที่อาศัยแต่เพียงความรู้ความจำเท่านั้น เป็นต้น

ข. ปัจจัยที่มีผลต่อความเที่ยง

ปัจจัยที่มีผลต่อความเที่ยงของแบบสอบมีหลายอย่าง เช่น ความยาวของแบบสอบ เวลาในการทำแบบสอบ ความเป็นเอกพันธ์ของกลุ่มตัวอย่าง ความยากของแบบสอบ ความเป็นปรนัย การกระจายของคะแนนสอบ และวิธีในการคำนวณหาค่าความเที่ยง เป็นต้น (Gronlund, ๑๙๗๖:๑๑๗-๑๒๒; Mehrens and Lehmann, ๑๙๗๕:๑๐๐-๑๐๓) ปัจจัยดังกล่าวนี้มีอิทธิพลต่อความเที่ยงดังนี้

๑. ความยาวของแบบสอบ

เมื่อแบบสอบมีความยาวเพิ่มขึ้น ก็จะมีค่าความเที่ยงเพิ่มขึ้น ซึ่งอาจแสดงจากสูตรต่อไปนี้

$$r_{tt} = \frac{kr}{1+(k-1)r}$$

ในเมื่อ r_{tt} = ค่าความเที่ยงที่คาดหวัง เมื่อแบบสอบยาวขึ้น k เท่า

k = อัตราส่วนของข้อสอบใหม่ที่เพิ่มขึ้น

r = ค่าความเที่ยงเดิม

สมมติว่า แบบสอบชุดหนึ่งมีค่าความเที่ยง = .๖๐ ถ้าเพิ่มข้อสอบที่วัดสิ่งเดียวกัน และมีลักษณะคล้ายคลึงกันอีก ๓ เท่า แบบสอบนั้นจะมีค่าความเที่ยงใหม่ดังนี้

$$r_{tt} = \frac{3 \times .60}{1 + (2 \times .60)} = .818$$

๒. เวลาในการทำแบบสอบ

ถ้าเวลาในการทำแบบสอบไม่เหมาะสม จะมีผลต่อค่าความเที่ยงของแบบสอบ เช่น ถ้าเวลาน้อยเกินไป ผู้สอบทำไม่ทันก็อาจเดาคำตอบ ซึ่งจะทำให้การกระจายของคะแนนแตกต่างจากแบบสอบที่ให้เวลาเพียงพอ แต่ถ้าหากว่าแบบสอบให้เวลาผู้สอบมากเกินไปจะทำให้คนที่เรียนเก่งและเรียนอ่อนทำแบบสอบได้คะแนนเท่า ๆ กัน คะแนนจะมีการกระจายน้อย และมีผลต่อความเที่ยงของแบบสอบเช่นกัน ในปี ๑๙๓๒ Cook (อ้างจาก Hopkins, ๑๙๖๔:๒๗๑) พบว่า เวลาในการสอบมีผลต่อค่าความเที่ยงอย่างมีนัยสำคัญ กล่าวคือ แบบสอบการสะกดคำฉบับหนึ่ง ถ้าให้เวลาพอดีจะมีค่าความเที่ยง .๘๑ - .๘๒ แต่ถ้าให้เวลามากเกินไป ความเที่ยงจะลดลงเป็น .๗๑ - .๘๕ เท่านั้น

๓. ความเป็นเอกพันธ์ของกลุ่มตัวอย่าง

ถ้าปัจจัยอย่างอื่นที่มีผลต่อแบบสอบ ๒ ฉบับเท่าเทียมกัน แบบสอบที่สอบกับกลุ่มตัวอย่างวิธีหนึ่งมีค่าความเที่ยงมากเท่าใด ก็ยังมีค่าความเที่ยงมากขึ้นเท่านั้น ทั้งนี้จะสังเกตได้จากสูตรทั่วไปของค่าความเที่ยงดังนี้

$$r_{tt} = 1 - \frac{\sigma_e^2}{\sigma_t^2}$$

ในเมื่อ σ_e^2 = ความแปรปรวนของความคลาดเคลื่อน

σ_t^2 = ความแปรปรวนของคะแนนสอบทั้งหมด

เนื่องจาก σ_e^2 เป็นความแปรปรวนของคะแนนที่ปรากฏ (observed score) ของกลุ่มตัวอย่างเกี่ยวกับคะแนนจริง (true score) ว่ามีความคลาดเคลื่อนเล็กน้อยเพียงใด ดังนั้น σ_e^2 จึงเป็นค่าที่คงที่ แม้ว่ากลุ่มตัวอย่างจะเปลี่ยนไป แต่ค่า σ_t^2 จะมีความเพิ่มขึ้นเมื่อกลุ่มตัวอย่างมีความแตกต่างกันมากขึ้น จึงทำให้ค่าความเที่ยงมีมากขึ้น

ดังนั้น ถ้าแบบสอบนั้นใช้สอบกับกลุ่มเอกพันธ์ เช่น ใช้กับกลุ่มตัวอย่างที่คัดสรรแล้ว เช่น กลุ่มตัวอย่างที่สอบผ่านเข้ามหาวิทยาลัยได้แล้วจากคณะใดคณะหนึ่ง ค่าความเที่ยงของแบบสอบบจะลดลง (Cronbach, ๑๙๔๙:๑๓๓)

๔. ความสัมพันธ์ซึ่งกันและกันของข้อสอบ

แบบสอบใดที่ข้อสอบมีความสัมพันธ์ซึ่งกันและกันสูง แบบสอบนั้นจะมีค่าความเที่ยงสูง (Guilford, ๑๙๕๔:๓๖๐) ในปี ๑๙๗๑ Bowers (Bowers, ๑๙๗๑: ๔๒๗) ได้เสนอสูตรในการคำนวณค่าความเที่ยงโดยอาศัยค่าสหสัมพันธ์ซึ่งกันและกันของข้อสอบดังนี้

$$r_{tt} = \frac{n\bar{r}_{ij}}{1+(n-1)\bar{r}_{ij}}$$

ในเมื่อ \bar{r}_{ij} = ค่าเฉลี่ยของค่าสัมประสิทธิ์สหสัมพันธ์ซึ่งกันและกันของข้อสอบ

n = จำนวนข้อสอบ

นอกจากนี้ แบบสอบที่ข้อสอบมีความสัมพันธ์กับคะแนนรวมยิ่งมาก แบบสอบนั้นก็ยังมีค่าความเที่ยงมาก ในปี ๑๙๖๔ Gaylord (Gaylord, ๑๙๖๔:๓๐๓) ได้เสนอสูตรในการคำนวณหาค่าความเที่ยง โดยอาศัยค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างข้อสอบกับคะแนนรวมไว้ดังต่อไปนี้

$$\bar{r}_{tt} = \frac{n\bar{r}_{it}^2 - 1}{(n-1)\bar{r}_{it}^2}$$

ในเมื่อ n = จำนวนข้อสอบ

\bar{r}_{it} = ค่าเฉลี่ยของค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างข้อสอบกับคะแนนรวม เช่น ค่า r_{bis} เป็นต้น

นอกจากนี้ Gaylord ได้พิสูจน์ว่า สูตรที่ Guilford เสนอไว้เพื่อการคำนวณหาค่าความเที่ยงจากค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างข้อสอบกับคะแนนรวมนั้นไม่ถูกต้องคือสูตรต่อไปนี้

$$r_{tt} = \frac{n\bar{r}_{it}^2}{1+(n-1)\bar{r}_{it}^2}$$

๕. ความยากง่ายของข้อสอบ

ข้อสอบที่ยากเกินไปหรือง่ายเกินไปไม่มีอำนาจจำแนก แบบสอบที่มีข้อสอบที่ง่ายเกินไปหรือยากเกินไปจำนวนมาก จะมีค่าความเที่ยงต่ำ Gronlund (Gronlund, ๑๙๗๖:๑๒๑) แนะนำว่า แบบสอบเพื่อการคัดเลือกที่ดีควรมีค่าความยากโดยเฉลี่ยเท่ากับค่าจุดกึ่งกลางระหว่างค่าโอกาสการเดา กับจำนวนข้อสอบทั้งหมด เช่น ถ้าข้อสอบแบบเลือกตอบมี ๔ ตัวเลือก ยาว ๕๐ ข้อ ความยากง่ายที่เหมาะสมควรเป็น $[(๒๕-๕๐)/๒] + ๒๕ = ๓๘$ (หรือ $p = .๓๘$) เป็นต้น ทั้งนี้เพื่อให้ได้แบบสอบที่มีค่าความเที่ยงสูง Cureton (Cureton, ๑๙๖๖:๑๕) เสนอแนะว่า แบบสอบที่มีข้อสอบง่าย ๆ อยู่ไม่ควรหาค่าความเที่ยงโดยวิธี KR_{21} เพราะค่าที่คำนวณได้จะต่ำกว่าความเป็นจริง ทั้งนี้เพราะว่าค่าความแปรปรวน (σ_t^2) จะคงที่ แต่ค่าความแปรปรวนรายข้อเพิ่มขึ้น ค่าจึงต่ำกว่าความเป็นจริง

๖. ความเป็นปรนัยในการให้คะแนน

แบบสอบที่ให้คะแนนไม่เป็นปรนัยค่าความเที่ยงจะไม่แน่นอน ดังนั้นควรทำเครื่องมือในการวัดให้สามารถให้คะแนนมีลักษณะเป็นปรนัยให้ได้มากที่สุดเท่าที่จะทำได้

๗. วิธีการคิดค่าความเที่ยง

การคิดค่าความเที่ยงมีหลายวิธี แต่ละวิธีให้ผลต่างกัน และแต่ละวิธีก็เหมาะสมสำหรับจุดมุ่งหมายเฉพาะอย่างในการตีความของค่าความเที่ยง ค่าความเที่ยงที่คิดแบบแบ่งครึ่ง (Split-half method) มักให้ค่าสูงที่สุด ค่าความเที่ยงที่คิดแบบการสอบซ้ำ

(Test-retest method) มักให้ค่าปานกลาง และค่าความเที่ยงที่คิดแบบ Kuder-Richardson มักให้ค่าต่ำ เป็นต้น (Gronlund, ๑๙๗๖:๑๒๒)

๘. การเดา

การเดามีผลต่อค่าความเที่ยงมาก เพราะการเดาเป็นแหล่งทำให้เกิดความคลาดเคลื่อนในการวัดมาก Magnusson (Magnusson, ๑๙๖๖:๑๐๑) กล่าวว่า แหล่งความคลาดเคลื่อนในการวัดที่สำคัญ มี ๓ แหล่งคือ การบริหารการสอบ การคิดคะแนน และการเดา ถ้ามีการเดามากแบบสอบจะมีค่าความเที่ยงลดลง Mattson (Mattson, ๑๙๖๔:๗๒๔) ได้เสนอสูตรในการทำนายค่าความเที่ยงของแบบสอบที่มีการเดาถูกดังนี้

$$\frac{r_{tt} p_c (1-p_c)}{1-r_{tt}} = n(1-\pi)^2 s_{p_k}^2$$

ในเมื่อ p_c = ค่าเฉลี่ยของกลุ่มผู้สอบที่ทำแบบสอบได้ถูกต้อง

r_{tt} = ค่าความเที่ยงเดิม

n = จำนวนข้อสอบ

π = โอกาสที่ผู้สอบทั้งกลุ่มเดาได้ถูกต้อง

$s_{p_k}^2$ = ค่าความแปรปรวนคะแนนที่ทำถูกโดยไม่ได้เดา

ดังนั้น ถ้าแบบสอบชุดหนึ่งมี ๑๐๐ ข้อ ผู้สอบโดยเฉลี่ยแล้วทำถูกร้อยละ ๕๐ ค่าความแปรปรวนของคะแนนที่ผู้สอบทำถูกโดยไม่ได้เดา = .๐๑ แบบสอบชุดนี้จะมีค่าความเที่ยง = .๕๐ แต่ถ้าแบบสอบชุดเดียวกันนี้ มีโอกาสเดาได้ถูกต้องร้อยละ ๕๐ ค่าเฉลี่ยของกลุ่มผู้สอบจะทำข้อสอบได้ถูกร้อยละ ๗๕ และความแปรปรวนของคะแนนที่ทำถูก = .๑๐ แล้ว แบบสอบชุดเดิมนี้อาจมีค่าความเที่ยงเพียง .๕๗๑ เท่านั้น

๙. คำสั่งในการทำแบบสอบ

คำสั่งในการทำแบบสอบเป็นตัวแปรที่สำคัญมากในการทำให้แบบสอบมีค่าความเที่ยงเปลี่ยนแปลงไปได้ โดยเฉพาะอย่างยิ่งคำสั่งที่ให้ผู้สอบเดาข้อสอบที่ทำได้จากการศึกษาของ Sax และ Collet (Sax and Collet, ๑๙๖๘:๑๓๔) ให้กลุ่มตัวอย่าง ๒๔๐ คน แบ่งเป็น ๓ กลุ่มเท่ากันแล้ว ให้กลุ่มตัวอย่างทำแบบสอบ Nelson Test of Mental Ability ซึ่งยาว ๕๐ ข้อ คำสั่งมี ๓ อย่างคือ ๑) บอกให้ผู้สอบเดาคำตอบข้อสอบที่ทำได้ และไม่หักคะแนนการเดา ๒) ห้ามผู้สอบที่ทำได้เดาคำตอบข้อสอบถ้าเดาผิดจะหักข้อละ ๑ คะแนน (แบบสอบมี ๔ ตัวเลือก) และ ๓) ห้ามผู้สอบที่ทำได้เดา แต่ถ้าบอกว่าตัวเลือกใดบ้างที่ผิดจะได้หัก ๑ คะแนน ผลจากการวิจัยพบว่า ค่าความเที่ยงของแบบสอบที่มีคำสั่งอย่าง ๒ และ ๓ เพิ่มขึ้นจากเดิมอย่างมีนัยสำคัญ นอกจากนี้จากการศึกษาของ Traub และ Hambleton (Traub and Hambleton, ๑๙๗๒:๗๓๗) พบว่า คำสั่งในการทำแบบสอบ ๒ อย่างคือ ๑) ห้ามไม่ให้ผู้สอบเดาคำตอบที่ไม่รู้ และให้รางวัลบางอย่างสำหรับข้อที่ทำได้แล้วเว้นไว้ และ ๒) ให้เดาคำตอบได้แต่ถ้าผิดจะหักคะแนน ผลปรากฏว่าแบบสอบศัพท์ยาว ๕๐ ข้อ ที่ใช้คำสั่งที่ ๑ มีค่าความเที่ยงเพิ่มขึ้นจากแบบสอบที่ใช้คำสั่งที่ ๒ อย่างมีนัยสำคัญ ต่อมาเขาได้เพิ่มคำสั่งที่ ๓ ขึ้นอีกคือ ให้ผู้สอบทำข้อสอบทุกข้อ หากทำไม่ได้ให้เดาคำตอบ จากการศึกษากับกลุ่มตัวอย่าง ๑,๐๔๑ คน โดยใช้แบบสอบคณิตศาสตร์ และแบบสอบศัพท์ยาว ๒๐ ข้อ เป็นเครื่องมือ ผลปรากฏว่า แบบสอบที่ใช้คำสั่งที่ ๑ มีค่าความเที่ยงสูงกว่าแบบสอบที่ใช้คำสั่งที่ ๒ และ ๓ อย่างมีนัยสำคัญ แบบสอบที่ใช้คำสั่งที่ ๓ มีค่าความเที่ยงต่ำที่สุด

๑๐. การคิดค่าน้ำหนักคะแนนของแบบสอบ

การให้ค่าน้ำหนักคะแนนของข้อสอบรายข้อและรายตัวเลือก เป็นวิธีหนึ่ง ที่ตามทฤษฎีแล้วเชื่อว่า จะทำให้แบบสอบมีคะแนนการกระจายเพิ่มมากขึ้นกว่าการให้คะแนนวิธีประเพณีนิยม (Wang and Stanley, ๑๙๗๐:๖๖๓) ผลการวิจัยเกี่ยวกับการให้ค่าน้ำหนักคะแนนรายข้อและรายตัวเลือก ได้กล่าวไว้แล้วในตอน ๑ ของบทที่ ๒ นี้

๑๑. ปัจจัยอื่น ๆ ที่มีผลต่อความเที่ยงปรากฏคู่ขนาน (Parallel Extrinsic Reliability)

Hopkins (Hopkins, ๑๙๖๔:๒๗๑-๒๘๑) ได้ทำการศึกษาปัจจัยต่าง ๆ ที่มีผลต่อความเที่ยงของแบบสอบ ซึ่งยังไม่มีวิธีการวัดได้แน่นอน แต่มีความสำคัญต่อการสอบ เขาเรียกความเที่ยงนี้ว่า "ความเที่ยงปรากฏคู่ขนาน" ปัจจัยต่าง ๆ นี้ได้แก่

๑. ความเร็วในการทำแบบสอบ
๒. การเดา
๓. รูปแบบ (format) ของแบบสอบ
๔. เวลาที่กำหนดให้ทำแบบสอบ
๕. กิจนิสัยในการทำแบบสอบ (response style/response set)
๖. คำสั่งในการทำแบบสอบ
๗. การทำข้อสอบไม่ครบทุกข้อ
๘. ระยะเวลาในการสอบซ้ำ
๙. เวลาหรือโอกาสในการทำแบบสอบ
๑๐. ความเหน็ดเหนื่อยและความเบื่อหน่ายในการทำแบบสอบ

Hopkins กล่าวว่า แบบสอบที่มีค่าความเที่ยงซึ่งวัดโดยวิธีประเพณีนิยม (Conventional Reliability) เท่ากัน อาจมีความเที่ยงปรากฏคู่ขนานแตกต่างกันก็ได้

ตอนที่ ๔ มโนทัศน์และงานวิจัยเกี่ยวกับการอ่านเข้าใจความ

ก. คำนิยาม

เป็นการยากที่จะให้คำนิยามว่า การอ่านเข้าใจความคืออะไร เพราะว่ามีแนวคิดเกี่ยวกับเรื่องดังกล่าวมากมาย เนื่องจากการอ่านเข้าใจความเกี่ยวข้องกับศาสตร์หลายสาขา เช่น จิตวิทยา ภาษาศาสตร์ จิตภาษาศาสตร์ สังคมศาสตร์ สังคมภาษาศาสตร์ พลศาสตร์และครุศาสตร์ ดังนั้น แนวคิดของผู้รู้ในแต่ละศาสตร์จึงแตกต่างกัน แต่ก็พอจะจำแนกความคิดออกได้เป็น ๓ แนวคิดใหญ่ ๆ (Harris, ๑๙๖๐:๑๐๗๕) คือ

๑. "การอ่านเข้าใจความ คือ ขบวนการรับรู้คำ และความหมายของคำทางสายตา" ดังนั้น การศึกษาการอ่านจึงมุ่งเน้นที่การเคลื่อนไหวของสายตา ระยะเวลาในการอ่าน และการพัฒนาช่วงของการจำ เป็นต้น ผู้นำด้านแนวคิดดังกล่าว ได้แก่ Vernon, Thinker และ Buswell

๒. "การอ่านเข้าใจความ คือ ขบวนการขยายความหมายของคำหรือการคิดที่สัมพันธ์กับสัญลักษณ์ที่เขียนขึ้น" แนวความคิดนี้เน้นที่การเข้าใจการสื่อสาร เช่น การแปล การย่อความ และการอ้างอิงอารมณ์ผู้เขียน เป็นต้น ผู้นำแนวคิดนี้ ได้แก่ Thorndike, Gray และ Richards

๓. "การอ่าน คือ ขบวนการของการรับรู้และเข้าใจสาระที่เขียนขึ้นในลักษณะที่ดูขนานกับการสื่อความหมายด้วยการพูด" แนวคิดนี้เน้นที่ขบวนการทางภาษาศาสตร์ จิตวิทยา และสังคมวิทยา ผู้นำของทัศนะแนวนี้ ได้แก่ Carroll และ Bloomfield เป็นต้น

นอกจากนี้ ยังมีผู้ให้คำนิยามและทฤษฎีของการอ่านไว้อีกมากมาย เช่น ทฤษฎีของ Holmes (อ้างจาก Karlin, ๑๙๗๑:๒๐) กล่าวว่า "การอ่านคือ ขบวนการของการคิด เนื่องจากสมองเก็บสาระไว้ ๓ รูปแบบ คือ ๑) สาระจากการมองเห็น ๒) สาระจากการได้ยิน และ ๓) สาระจากการรับรู้ทางอื่น สาระต่าง ๆ เหล่านี้ จะมีปฏิสัมพันธ์

กันอย่างมีระเบียบตามลำดับชั้น ทำให้เกิดความสามารถของการอ่าน แต่ Erdman และคณะ (อ้างจาก Stauffer, ๑๙๗๕:๖) เชื่อว่า การอ่านเข้าใจความคือ ขบวนการทางสมองที่มีการเปลี่ยนแปลง และตื่นตัวอยู่เสมอ และเข้าใจคำที่เขียนเป็นคำ ๆ หรือกลุ่มคำว่า หมายความว่าอย่างไร

แต่อย่างไรก็ตาม ในปัจจุบันนี้มีความเชื่อมั่นว่า การอ่านเป็นขบวนการทางจิตสังคมภาษาศาสตร์ ที่มีบูรณาการผสมกลมกลืนกัน ที่หลายคน เช่น Bloom (๑๙๕๖) และ Gagne (๑๙๗๐) เชื่อว่ามีการเรียนรู้เป็นลำดับชั้น (hierarchies)

ข. ลำดับชั้นความเข้าใจในการอ่าน

การวิจัยเกี่ยวกับลำดับชั้นความเข้าใจในการอ่านมีมาก และได้ข้อสรุปค่อนข้างแน่ว่า การอ่านเข้าใจความมีลำดับชั้น เพราะได้มีการตรวจสอบกันหลายวิธี เช่น การวิเคราะห์กระสวนการตอบ การวิเคราะห์วิธีสัมพันธ์ และทฤษฎีลำดับชั้น (Bergan, ๑๙๘๐:๖๒๕; Airasian, ๑๙๗๕:๑๖๓) เป็นต้น แต่ว่ายังไม่มียผลงานวิจัยที่แน่ชัดว่า ลำดับชั้นความเข้าใจในการอ่าน และลำดับของการสอน (teaching order) เพื่อให้เกิดความเข้าใจมีความสัมพันธ์กันมากนักน้อยเพียงใด (Bergan, ๑๙๘๐:๖๔๓)

ในปี ๑๙๘๔ Hill (Hill, ๑๙๘๔:๒๓๑) ได้ทำการวิจัยเกี่ยวกับลำดับชั้นความเข้าใจในการอ่านภาษาอังกฤษ โดยใช้ LISREL Method (Linear Structural Relationships) ซึ่งเป็นสถิติขั้นสูง เพื่อการศึกษาและตรวจสอบความสัมพันธ์ของตัวประกอบ เขาพบว่า ในระดับมหภาค (macro-level) แล้ว การอ่านเข้าใจความมีเพียงมิติเดียว (unidimension) แต่ในระดับจุลภาค (micro-level) นั้น การอ่านเข้าใจความมีหลายตัวประกอบและมีลำดับชั้น การค้นพบดังกล่าวนี้สอดคล้องกับผลการวิจัยของ Andrich และ Godfrey ที่ทำในปี ๑๙๗๘ โดยการวิเคราะห์ข้อสอบด้วยรูปแบบบราสซ์แล้วจึงวิเคราะห์ตัวประกอบของข้อสอบ การวิจัยพบว่า ลำดับชั้นความเข้าใจในการอ่านเรียงจากต่ำที่สุดไปยังสูงที่สุด เป็นดังนี้ (อ้างจาก Hill, ๑๙๘๔:๒๒๙)

๑. การเข้าใจเนื้อเรื่องที่อ่านที่ระบุไว้อย่างชัดเจน
๒. การเข้าใจเนื้อเรื่องที่ต้องการสรุปเนื้อหาของข้อความ
๓. การสรุปเรื่องที่อ่าน
๔. การวินิจฉัยความหมายของศัพท์จากเนื้อเรื่อง
๕. การจำความหมายของศัพท์
๖. การติดตามโครงเรื่องของเนื้อหาที่อ่าน
๗. การระบุวิธีการเขียนของผู้แต่ง
๘. การเข้าใจอารมณ์ แนวคิด และจุดมุ่งหมายของผู้แต่ง

ก. ตัวประกอบของการอ่านเข้าใจความ

จากการวิจัยของ Holmes (อ้างจาก Karlin, ๑๙๗๑:๒๐) ที่ทำการสอบนักเรียนระดับมัธยมศึกษาด้วยแบบสอบ ๕๐ ชุด พบว่า ความหมายของศัพท์ในประโยค และความหมายของศัพท์โดด ๆ มีส่วนในการอ่านเข้าใจความร้อยละ ๓๒ การอุปมาอุปไมยทางภาษา และการฟังในระหว่างการอ่านมีส่วนร้อยละ ๓๒ อีกร้อยละ ๒๕ เป็นความสามารถด้านอื่น ๆ แต่ต่อมา Singer (อ้างจาก Karlin, ๑๙๗๑:๒๐) ได้ทำการวิจัยเกี่ยวกับความสามารถในการอ่านเข้าใจความของนักเรียนในระดับ grade ๓-๖ พบว่า นักเรียนในระดับดังกล่าวนี้ ความสามารถในการจำคำ และความสัมพันธ์ระหว่างคำกับการมองเห็นมีส่วนในการเข้าใจความประมาณร้อยละ ๗๐ และอีกร้อยละ ๓๐ เป็นความสามารถจากด้านอื่น ๆ

แต่อย่างไรก็ตาม Fry (Fry, ๑๙๖๓: vii) มีความเห็นว่า การอ่านเข้าใจความประกอบด้วยส่วนสำคัญ ๒ ส่วนคือ

๑. การเข้าใจข้อเท็จจริงที่ระบุไว้โดยตรง ซึ่งเขาเชื่อว่ามีส่วนในการอ่านเข้าใจความประมาณร้อยละ ๕๐

๒. การเข้าใจธรรมชาติทั่ว ๆ ไป และการอ้างอิงโดยทางอ้อมของเนื้อเรื่อง ซึ่งประกอบด้วย ก) การอ้างอิงสรุป ข) การสรุปนัยของข้อความ ค) การเข้าใจอารมณ์ของผู้เขียน และ ง) ความสามารถในการใช้เหตุผลทั่ว ๆ ไป Fry เชื่อว่าส่วนประกอบนี้ทำให้เกิดความเข้าใจในการอ่านอีกประมาณร้อยละ ๔๐ นอกจากนั้นอีกประมาณร้อยละ ๑๐ เกิดจากสาเหตุอื่น ๆ

นอกจากนี้ ยังมีผลงานวิจัยที่น่าสนใจเกี่ยวกับตัวองค์ประกอบการอ่าน คือ งานของ McKillop (McKillop, ๑๙๕๓) ซึ่งได้ทำการสอบนักเรียนในระดับมัธยม ๔ แห่งที่มีลักษณะคล้ายคลึงกัน จำนวน ๕๑๒ คน เมื่อปี ๑๙๕๒ ด้วยเรื่องที่ให้กลุ่มตัวอย่างอ่าน ๒ ลักษณะ คือ เรื่องที่นักเรียนบอกว่าชอบ และเรื่องที่ไม่ชอบ แบบสอบที่ใช้เป็นแบบเลือกตอบที่ประกอบด้วยคำถาม ๔ ประเภท คือ ๑) ถามข้อเท็จจริง ๒) ถามการสรุปนัยข้อความ ๓) ถามการใช้วิจารณ์ และ ๔) ถามการประเมินจากข้อเท็จจริง

ผลของการวิจัยพบว่า

๑. นักเรียนที่มีเจตคติที่ดีต่อเรื่องที่อ่าน จะทำข้อสอบชนิด ๑ ๒ ๓ และ ๔ ได้ถูกต้องมากขึ้นตามลำดับอย่างมีนัยสำคัญ
๒. นักเรียนที่มีเจตคติที่ไม่ดีต่อเรื่องที่อ่าน จะทำข้อสอบชนิดที่ ๑ และ ๒ ได้คะแนนไม่แตกต่างจากนักเรียนที่มีเจตคติที่ดีต่อเรื่องนั้น ๆ แต่จะทำข้อสอบชนิดที่ ๓ และ ๔ ได้คะแนนน้อยกว่าอย่างมีนัยสำคัญ

ดังนั้น เราอาจสรุปได้ว่า การอ่านเข้าใจความวิชาวาษาอังกฤษนั้น มีลำดับชั้นความเข้าใจในการอ่านหลายระดับ ตามลักษณะตัวประกอบของการอ่านต่าง ๆ ดังได้กล่าวมาแล้ว และลำดับขั้นเหล่านั้นจะมีความสัมพันธ์กับระดับความยากง่ายของการเรียนรู้ด้วย กล่าวคือ ความเข้าใจในระดับต้น ๆ เช่น การอ้างอิงโดยตรง มีความง่ายที่สุด (McKillop, ๑๙๕๓: Fry, ๑๙๖๓:vii) ส่วนทักษะอื่น ๆ นั้น จะยากกว่าผู้อ่านที่ใช้ความรู้ความพยายามในการอ่านมากกว่า ก็น่าที่จะได้คะแนนในการสอบมากกว่า การสอบดังกล่าวนี้ น่าจะมีความ

ตรงเชิงภาวะสันนิฐาน (Construct Validity) มากกว่าการที่ให้คะแนนวิธีประเพณีนิยม
 ดังนั้น การสอบที่ให้คะแนนตัวเลือก หรือข้อสอบต่างกัน น่าจะมีความตรงและเที่ยงมากกว่า
 การให้คะแนนโดยวิธีประเพณีนิยมที่ใช้กันทั่วไป

ง. แนวคิดใหม่เกี่ยวกับการอ่านเข้าใจความ

ในปัจจุบันนี้เป็นที่ยอมรับกันทั่วไปแล้วว่า ความสามารถในการอ่านเข้าใจความ
 ข้อความต่าง ๆ นั้น ไม่ได้ขึ้นอยู่กับความสามารถของผู้อ่านด้านความรู้ทางภาษา
 ศาสตร์เท่านั้น แต่ยังขึ้นอยู่กับความรอบรู้ทั่ว ๆ ไป และความสามารถทางสมองในการ
 จัดกระทำกับความรู้ความสามารถดังกล่าวแล้วด้วย (Carrell, ๑๙๘๓:๑๘๓) และมี
 ผู้วิจัยบางท่าน เช่น Kintsch (๑๙๗๕) Rumelhart (๑๙๗๕) และ Thorndyke
 (๑๙๗๗) เป็นต้น (อ้างจาก Carrell, ๑๙๘๓:๑๘๓) พบว่า ผู้อ่านยังมีพื้นฐานความรู้
 เกี่ยวกับเรื่องที่อ่าน หรือเกี่ยวกับโครงสร้างความรู้ของเรื่องที่อ่านมากเท่าใด ก็จะมี
 สามารถอ่านเข้าใจเรื่องที่อ่านได้มากยิ่งขึ้น ดังนั้น Carrell (Carrell, ๑๙๘๔:๓๓๓)
 จึงแนะนำ วิธีที่จะทำให้ผู้อ่านเข้าใจเรื่องที่อ่านได้วิธีหนึ่งก็คือ การสอนศัพท์ที่ก่อให้เกิด
 โครงสร้างความรู้ของเรื่องที่อ่านนั้นด้วย

อนึ่ง ความคิดเกี่ยวกับโครงสร้างความรู้ของผู้อ่าน และของเรื่องที่
 อ่านว่า มีผลต่อความสามารถในการอ่านเข้าใจความนั้น มีมานานแล้ว เพราะคำว่าโครงสร้าง
 สร้างความรู้ Bartlett เป็นผู้นำมาใช้แล้วตั้งแต่ปี ๑๙๓๒ ซึ่งแนวคิดนี้พอสรุปได้ดังต่อไปนี้

ทฤษฎีโครงสร้างความรู้

คำว่า โครงสร้างความรู้นี้ Mandler (Mandler, ๑๙๗๕:๒๖๓) ได้ให้คำนิยาม
 ไว้ดังนี้

"โครงสร้างความรู้ คือ โครงสร้างของความรู้ที่ไม่ได้ประกอบด้วย
 สมาชิกที่อยู่ในกลุ่มเดียวกัน หรือคล้ายคลึงกัน แต่ว่าเป็นโครงสร้างที่สมาชิก
 ประกอบกันเชิงมิติสัมพันธ์ และ/หรือสัมพันธ์กันชั่วคราว ส่วนต่าง ๆ เหล่านี้

สัมพันธ์กันตามประสบการณ์ในอดีตของแต่ละบุคคล เกี่ยวกับสิ่งของ ฉากสภาพการณ์ หรือเหตุการณ์ต่าง ๆ รวมทั้งความคาดหวัง (ปกติ จะอยู่ในสภาพจิตไร้สำนึก) ว่าสิ่งต่าง ๆ เหล่านี้เคยเกิดขึ้นอยู่ใน สภาพใด ส่วนโครงสร้างความรู้ย่อยมีสภาพเป็นตัวแปร หรือ "ช่องว่าง" ที่อาจจะ "บรรจุ" หรือ "เกิดมี" สิ่งที่เคยเกี่ยวข้อง ด้วยมาแล้วในอดีตเร็วหรือช้า แล้วแต่โอกาสของสิ่งเหล่านี้ว่า เคย เกิดขึ้นใหม่น้อยเพียงใด"

อนึ่ง Rumelhart และ Ortory (อ้างจาก Rumelhart, ๑๙๘๐:๔๐) ได้สรุป คุณสมบัติเฉพาะของโครงสร้างความรู้ไว้ว่า มี ๖ อย่าง คือ

๑. โครงสร้างความรู้ประกอบด้วยตัวแปรต่าง ๆ
๒. โครงสร้างความรู้หนึ่งสามารถที่จะแฝงอยู่ในโครงสร้างความรู้อื่นได้
๓. โครงสร้างความรู้เป็นตัวแทนของความรู้ทางนามธรรมทุก ๆ ระดับได้
๔. โครงสร้างความรู้เป็นตัวแทนของความรู้มากกว่าจะเป็นคำนิยามของสิ่งใด ๆ
๕. โครงสร้างความรู้เป็นขบวนการที่มีกิจกรรมอยู่เสมอ ๆ
๖. โครงสร้างความรู้เป็นเครื่องมือแสวงหาความรู้ ซึ่งมีกระบวนการที่มุ่งที่จะ ประเมินความสอดคล้องของข้อมูลที่กำลังจัดกระทำอยู่

นอกจากนี้ Rumelhart (Rumelhart, ๑๙๘๐:๓๔) ยังกล่าวว่า โครงสร้าง ความรู้นี้ มีลักษณะเหมือนกับโปรแกรมคอมพิวเตอร์ เพราะสามารถประเมินได้ว่า ข้อมูลที่พบ ในการอ่านนั้น เหมาะสมกับโครงสร้างเดิมของผู้อ่านหรือไม่ ดังนั้น เขาจึงสรุปว่า ทฤษฎี โครงสร้างความรู้ี้ มีลักษณะเป็นทั้งทฤษฎีสาระข้อความ (Prototype Theory) และทฤษฎี วิธีกร (Procedural Theory) และ Mandler (Mandler, ๑๙๗๔:๒๖๓-๖) กล่าวว่า โครงสร้างความรู้ที่มีความสำคัญในการทำให้เกิดความเข้าใจในการสื่อสารที่สำคัญนั้นมี ๒ อย่าง คือ

๑. โครงสร้างความรู้เกี่ยวกับฉากสภาพการณ์ ได้แก่ โครงสร้างของความคิดที่บุคคลคาดหวังว่า จะได้พบเห็นสิ่งใดบ้างในฉากสภาพการณ์ต่าง ๆ เช่น โครงสร้างความรู้เกี่ยวกับห้อง ก็อาจมีผนังห้อง เก้าอี้ หน้าต่าง และเพดาน เป็นต้น สิ่งต่าง ๆ เหล่านี้มีสภาพความเป็นอยู่เกี่ยวข้องกันเชิงมิติสัมพันธ์ในลักษณะ เฉพาะตามที่บุคคล เคยประสบมาก่อน

๒. โครงสร้างความรู้เกี่ยวกับเหตุการณ์ ได้แก่ โครงสร้างของความรู้ที่เกิดขึ้นในเหตุการณ์ต่าง ๆ ตามที่บุคคลเคยพบเห็น หรือภาคว่า น่าจะเกิดขึ้น

ดังนั้น จึงเป็นที่เชื่อกันว่า โครงสร้างของ เรื่องจะสะท้อนให้เห็นโครงสร้างความรู้ของผู้เขียน (อ้างจาก C. Duangploy, ๑๙๘๒:๒๔) ถ้าผู้อ่านมีโครงสร้างความรู้สอดคล้องกับโครงสร้างความรู้ของผู้เขียนหรือของ เรื่องแล้ว จะทำให้สามารถเข้าใจเรื่อง ที่อ่านได้ง่ายขึ้น แต่ถ้าหากว่าโครงสร้างความรู้ของผู้อ่านแตกต่างจากโครงสร้างของ เรื่อง หรือโครงสร้างความรู้ของผู้เขียนแล้ว การอ่านก็จะมีปัญหา

นอกจากนี้ เป็นที่เชื่อกันว่า โครงสร้างของ เรื่องมีลักษณะแทนโครงสร้างของ ความจำเรื่องนั้น ๆ ได้ และโครงสร้างนี้เป็นอิสระจากความซับซ้อนของโครงสร้างผิว (surface structure complexity) ของเรื่อง ที่อ่าน และสาระเก่า-ใหม่ของเรื่อง (proposition) ที่จำไว้นั้น เรียงกันตามลำดับชั้น (Kintsch and Monk, ๑๙๗๗; อ้างจาก C. Duangploy, ๑๙๘๒:๒๔) ด้วยเหตุนี้ จึงเชื่อกันว่า สาระของเรื่อง ที่ผู้อ่าน ระลึกได้ สามารถที่จะแปลงให้เป็นความสัมพันธ์ของสาระเก่า-ใหม่ได้ และเปรียบเทียบ กับโครงสร้างของ เรื่องที่อ่านได้ ดังนั้น การวัดความสามารถในการอ่านเข้าใจความด้วย การเปรียบเทียบดังกล่าวแล้ว จึงเป็นวิธีที่กำลังได้รับความนิยมมากในปัจจุบันว่า เป็นวิธี ที่เหมาะสมมากในการวัดความสามารถศักยภาพ (Competence) ในการอ่านเข้าใจความ เช่น ในงานวิจัยของ Carrell (Carrell, ๑๙๘๓) และของ Duangploy (C. Duangploy: ๑๙๘๒) เป็นต้น

อนึ่ง Rumelhart (Rumelhart, ๑๙๘๐:๔๑) ยังกล่าวว่า ชื่อเรื่องและรูปภาพประกอบเรื่อง ที่บุคคลอ่านเป็นโครงสร้างความรู้อย่างหนึ่งที่กระตุ้นให้ผู้อ่านเกิดมีโครงสร้างความรู้ย่อย ที่จะทำให้การอ่านเข้าใจความได้ในลักษณะ "เบื้องสูง-สูง-เบื้องต่ำ" แต่คำศัพท์บางคำในเรื่องที่มีความหมายสำคัญ จะมีสภาพเป็นโครงสร้างความรู้ย่อยที่สามารถทำให้ผู้อ่านเข้าใจเรื่องทีอ่านได้ในลักษณะ "เบื้องต่ำ-สูง-เบื้องสูง"

เมื่อไม่นานมานี้ ได้มีงานวิจัยเกี่ยวกับอิทธิพลของโครงสร้างความรู้ที่มีต่อการอ่านเข้าใจความภาษาอังกฤษที่น่าสนใจ ๒ เรื่อง กล่าวคือ

ในปี ๑๙๘๓ Carrell (Carrell, ๑๙๘๓) ได้ทำการวิจัยเกี่ยวกับผลกระทบของการให้ผู้อ่านรู้และไม่รู้สภาวะแวดล้อมของเรื่องทีอ่าน การมีคำศัพท์ที่สำคัญและไม่สำคัญในเรื่องทีอ่าน และความคุ้นเคยหรือความใหม่ของเรื่องทีอ่าน ผลปรากฏว่า

๑. สำหรับผู้อ่านที่เป็นเจ้าของภาษา สภาวะแวดล้อมของเรื่อง (ชื่อเรื่องและภาพประกอบเรื่อง) ศัพท์ที่สำคัญในเรื่อง และความใหม่ของเรื่องทีอ่าน มีผลต่อสัมฤทธิ์ผลในการอ่านอย่างมีนัยสำคัญ

๒. สำหรับผู้อ่านที่ไม่ใช่เจ้าของภาษา และมีความสามารถในการอ่านระดับเหนือปานกลางเล็กน้อย ศัพท์ที่สำคัญในเรื่องเท่านั้นที่มีอิทธิพลต่อสัมฤทธิ์ผลในการอ่านอย่างมีนัยสำคัญ

๓. สำหรับผู้อ่านที่ไม่ใช่เจ้าของภาษา แต่มีความสามารถในการอ่านในระดับสูง ความใหม่ของเรื่อง เท่านั้นที่มีอิทธิพลต่อสัมฤทธิ์ผลในการอ่านอย่างมีนัยสำคัญ

ดังนั้น จึงเป็นที่น่าสังเกตว่า สภาวะแวดล้อมของเรื่องและความคุ้นเคยกับเรื่องทีอ่านไม่ได้ทำให้ผู้อ่านที่ไม่ใช่เจ้าของภาษา (อังกฤษ) สามารถอ่านเรื่องทีอ่านได้เข้าใจความมากกว่าเรื่องทีไม่มีสภาวะแวดล้อมให้ หรือเรื่องทีใหม่เลย ไม่ว่าความสามารถในการอ่านของผู้อ่านจะอยู่ในระดับสูงหรือปานกลางก็ตาม แต่ที่สำคัญต่อผู้อ่านมากกว่า คือ คำศัพท์ที่สื่อความหมายสำคัญในเรื่อง

ต่อมาในปี ๑๙๘๔ Carrell (Carrell, ๑๙๘๔b:๘๖-๑๐๕) ได้ทำการวิจัยเกี่ยวกับอิทธิพลของโครงสร้างความรู้ต่อการอ่านเข้าใจความภาษาอังกฤษเป็นภาษาที่สองอีก โดยใช้กลุ่มตัวอย่างประชากรที่มีระดับความรู้ภาษาอังกฤษในระดับปานกลาง จำนวน ๔๐ คน และเรื่องสำหรับอ่าน ๓ เรื่อง ที่จัดเรียงเนื้อหาที่สำคัญต่างกัน ๒ แบบ โดยแบบหนึ่งเรียงเนื้อหาตามที่รู้จักทั่วไป แต่อีกแบบหนึ่ง เรียงเนื้อหาต่างออกไปจากที่รู้จักกันทั่วไป เมื่อให้ผู้อ่านเขียนรายละเอียดของเรื่องที่อ่านทั้ง ๒ แบบ ในเวลาที่ต่างกัน ๒๔ ช.ม. ผลปรากฏว่า ผู้อ่านสามารถให้สาระของเรื่องที่อ่านทั้ง ๒ แบบได้ไม่ต่างกันอย่างมีนัยสำคัญ แต่ให้อันดับที่ของสาระสำคัญของเนื้อหาได้ต่างกันอย่างมีนัยสำคัญ โดยให้สาระตามอันดับที่ตามเรื่องที่รู้จักกันทั่ว ๆ ไปได้ถูกต้องมากกว่าเรื่องที่เรียงลำดับเนื้อหาต่างออกไป นี่แสดงว่าโครงสร้างความรู้มีอิทธิพลต่อการอ่านอย่างมีนัยสำคัญ นอกจากนี้ Carrell (Carrell, ๑๙๘๔a:๑๐๕) ให้ความเห็นว่า การอ่านแล้วไม่เข้าใจส่วนหนึ่งเกิดจากผู้อ่านไม่มีโครงสร้างความรู้ของเรื่องที่เหมาะสม และอาจเกิดจากโครงสร้างความรู้ของผู้อ่านกับเรื่องที่อ่านแตกต่างกัน