# การปฏิรูปแนวคิดทฤษฎีการทดสอบแบบดั้งเดิมสู่การพัฒนาทฤษฎีการทดสอบแนวใหม่
# The Reformation Concept of Classical Test Theory (CTT) to Development Modern Test Theory.

นายเจนรบ โกรธา
นักศึกษาปริญญาโท สาขาการวัดและประเมินผลการศึกษา
คณะศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น

## บทนำ

การจัดการศึกษามีความจำเป็นที่จะต้องกำหนดจุดมุ่งหมายที่มีความชัดเจน เพื่อช่วยในการกำหนดคุณลักษณะที่ต้องการของผู้เรียน เพื่อใช้เป็นแนวทางในการวางแผนเพื่อใช้การจัดการเรียนรู้ จุดมุ่งหมายยังมีส่วนช่วยในการกำหนดทิศทางของการศึกษา การจัดประสบการณ์การเรียนรู้หรือกิจกรรมการจัดการเรียนรู้ที่เหมาะสม การวัดและการประเมินผลจึงเข้ามามีบทบาทสำคัญในการตัดสินผลการเรียนรู้ของผู้เรียนว่าสำเร็จตรงตามจุดมุ่งหมายที่กำหนดไว้หรือไม่ รวมทั้งยังช่วยให้สารสนเทศเพื่อช่วยในการปรับปรุงละพัฒนาการเรียนรู้ของผู้เรียนและการจัดกิจกรรมการเรียนรู้ของผู้สอน เมื่อการวัดและการประเมินผลเข้ามามีบทบาทสำคัญ ทำให้ทฤษฎีการทดสอบเข้ามามีบทบาทสำคัญในการวัดและประเมินผล เนื่องจากการการวัดและประเมินผลทางการศึกษานั้นเป็นสิ่งที่อยู่ภายในไม่สามารถวัดออกมาได้โดยตรง เพื่อทำความเข้าใจในสิ่งที่ต้องการวัดตามคุณลักษณะที่มุ่งวัด โครงสร้างของการวัด และการพัฒนาเครื่องมือ จึงจำเป็นต้องอาศัยทฤษฎีการทดสอบทดสอบเข้ามาช่วย โดยทฤษฎีการทดสอบนั้นหากแบ่งประเภทเป็น 2 ประเภทหลักใหญ่ๆ ได้แก่ ทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory) และทฤษฎีการทดสอบแนวใหม่ (Modern Test Theory)

ทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory) จะเน้นไปทางการตรวจสอบความสัมพันธ์ระหว่างคะแนนที่สังเกตได้กับคะแนนที่แท้จริง และวิเคราะห์คุณภาพโดยส่วนรวมของข้อสอบและแบบสอบ โดยใช้การวิเคราะห์คุณภาพของข้อสอบและแบบทดสอบ โดยใช้โมเดลคะแนนจริงแบบดั้งเดิม การใช้ทฤษฎีความเที่ยงและความคลาดเคลื่อน การใช้ความตรงเพื่อตรวจสอบความตรงต่างๆ หลักการการสร้างแบบทดสอบ และใช้หลักการวิเคราะห์ข้อสอบ ทฤษฎีการทดสอบแบบดั้งเดิมยังมีข้อด้อยของข้อตกลงเบื้องต้นมีข้อจำกัดของฐานความเชื่อเกี่ยวกับคะแนนความคลาดเคลื่อนและค่าพารามิเตอร์ของข้อสอบและแบบทดสอบ จึงส่งผลให้มีนักทฤษฎีการทดสอบหลายท่านได้ก่อการปฏิรูปแนวคิดเกิดขึ้น สู่ทฤษฎีการทดสอบแนวใหม่ (Modern Test Theory) เพื่อคลายข้อตกลงเบื้องต้น และแก้ไขจุดอ่อนบางประการ

ทฤษฎีการทดสอบแนวใหม่ (Modern Test Theory) จะมุ่งเน้นการขยายแนวคิดของทฤษฎีการทดสอบแบบดั้งเดิมให้มีขอบเขตที่กว้างขวาง และมีความน่าเชื่อถือมากยิ่งขึ้น โดยปรับข้อตกลงให้มีความสมเหตุสมผล ตามความจริงมากขึ้น เกี่ยวกับความคลาดเคลื่อนของการวัด ยอมให้มีความคลาดเคลื่อนจากการวัดนั้นมีความแตกต่างกันออกไปตามสถานการณ์การวัด เพื่อศึกษาความเที่ยงทั่วไป ภายใต้เงื่อนไขของการวัดตามแนวของ Generalizability Theory: G-Theory รวมทั้งศึกษาความคลาดเคลื่อนจากการวัดที่มีความแตกต่างกันไปตามลักษณะของข้อสอบและแบบสอบ ขึ้นกับระดับความสามารถของผู้สอบ และของข้อสอบและแบบสอบ ทั้งยังพยายามวัดคุณลักษณะภายในหรือความสามารถที่แท้จริงของบุคคลตามแนว Item

Response Theory (IRT) นอกจากนี้ความคลาดเคลื่อนจากการวัดของบุคคลเองก็ส่งผลต่อการตอบข้อสอบมากกว่า 1 องค์ประกอบ ดังนั้นเพื่อวัดความสามารถของบุคคลจึงมี 2 องค์ประกอบขึ้นไป เพื่อให้ความสามารถหลายมิติของบุคคลสอดคล้องกับข้อมูลได้ดียิ่งขึ้นและส่งผลต่อความน่าจะเป็นของการตอบถูก ตามแนวของ Multidimensional Item Response Theory (MIRT). ทั้งทฤษฎีการทดสอบแบบดั้งเดิมทฤษฎีการทดสอบแนวใหม่ ต่างก็ต้องการวัดที่เน้นผลผลิต และพฤติกรรมที่เกิดจากการนำความรู้และการนำทักษะไปใช้ในบริบทหรือสถานการณ์ที่เป็นชีวิตประจำวันจริงและด้วยการใช้สถานการณ์ที่เลียนแบบชีวิตจริง

KEYWORDS: Classical Test Theory, Modern Test Theory

**พัฒนาการทฤษฎีการทดสอบ**

พัฒนาการของทฤษฎีการทดสอบนั้นมีความสัมพันธ์กันอย่างมากกับศาสตร์ทางด้านจิตวิทยา โดยเริ่มต้นจากนักจิตวิทยาเองที่ต้องการทดสอลและหาวิธีการแก้ปัญหาหารวัดต่างๆ เพื่อพัฒนาจนเป็นศาสตร์การวัดและตรวจสอบที่มีความน่าเชื่อถืออย่างที่เราพบเห็นและใช้กันในปัจจุบัน

ศาสตร์ทางด้านการวัดและประเมินผลนั้นมีมายาวนาน แต่ตอนปลายศตวรรษที่ 19 ที่นักวิทยาศาสตร์และนักสถิติได้ให้ความสำคัญ มีการพัฒนาทฤษฎีการทดสอบอย่างมากและมีความรวดเร็วและกว้างขวาง รวมทั้งมีความเป็นระบบมาก เนื่องมาจากการรวมตัวกันของนักวัดผลเพื่อแลกเปลี่ยนความคิดผลงานวิชาการกันจนทำให้เกิดเป็นแนวคิดใหม่ๆ ทฤษฎีการทดสอบอย่างมากมายจนเป็นที่ยอมรับและสนใจอย่างกว้างขวาง

**ทฤษฎีการทดสอบแบบดั้งเดิม (CTT)** ทฤษฎีการทดสอบแบบดั้งเดิมเป็นทฤษฎีที่ถือปฏิบัติใช้กันมานานแล้ว ส่วนใหญ่จะรู้จักกันในชื่อ *ทฤษฎีคะแนนจริงแบบดั้งเดิม* ซึ่งองค์ความรู้ทั่วไปจะเน้นเกี่ยวกับการตรวจสอบความสัมพันธ์ระหว่างคะแนนที่สังเกตได้กับคะแนนจริง รวมทั้งวิเคราะห์คุณภาพโดยส่วนรวมของข้อสอบและแบบสแบสำหรับผู้สอบเฉพาะกลุ่ม

ทฤษฎีคะแนนจริงแบบดั้งเดิมนี้จะตั้งอยู่บนพื้นฐานของ โมเดลการวัดและข้อตกลงเบื้องต้นที่สำคัญหลายประการ โดยทฤษฎีนี้หลักๆ จะเน้นเพื่อการพัฒนาหลักการสร้างข้อสอบ การวิเคราะห์คุณภาพของแบบสอบ และการวิเคราะห์คุณภาพของข้อสอบ โมเดลคะแนนจริงแบบดั้งเดิม (X) เกิดจากองค์ประกอบที่สังเกตไม่ได้ 2 ส่วน คือคะแนนจริง (T) และคะแนนความคลาดเคลื่อน (E) ซึ่งมีความสัมพันธ์ดังสมการ

$$X = T + E$$

ซึ่งโมเดลการวัดตั้งอยู่บนข้อตกลงเบื้องต้น (Model Assumption) จำนวน 6 ข้อ ดังนี้

**ข้อตกลงเบื้องต้นข้อที่ 1** คะแนนที่ได้จากการวัด มีความสัมพันธ์แบบเส้นตรงและเชิงบวกกับคะแนนจริงและคะแนนความคลาดเคลื่อน

**ข้อตกลงเบื้องต้นข้อที่ 2** คะแนนจริงมีสภาวะคงที่ ซึ่งเท่ากับค่าเฉลี่ยของคะแนนที่ได้จากการวัดซ้ำๆ หลายๆครั้ง

**ข้อตกลงเบื้องต้นข้อที่ 3** คะแนนความ คลาดเคลื่อนไม่มีความสัมพันธ์กับคะแนนจริง

**ข้อตกลงเบื้องต้นข้อที่ 4** ความคลาดเคลื่อนส่วนบุคคล ไม่มีความสัมพันธ์ซึ่งกันและกัน

**ข้อตกลงเบื้องต้นข้อที่ 5** การทดสอบ 2 ฉบับจะเป็นแบบทดสอบคู่ขนาน เมื่อคะแนนจริงของผู้สอบมีค่าเท่ากันทั้งสองฉบับ

**ข้อตกลงเบื้องต้นข้อที่ 6** แบบทดสอบ 2 ฉบับ จะถือว่าเป็นแบบทดสอบทัดเทียมกัน

**ความเที่ยง (Reliability)**

เมื่อใดก็ตามที่มีการนำแบบทดสอบไปใช้ แบบทดสอบจำเป็นต้องมีคุณลักษณะสำคัญอีกประการที่เรียกว่า ความเที่ยง ซึ่งความเที่ยงนี้มีอยู่หลายประเภท การประมาณค่าสัมประสิทธิ์ความเที่ยงของแบบทดสอบเองก็สามารถทำได้หลายวิธีเช่นเดียวกัน

ทฤษฎีความเที่ยงนั้นก็ได้ประยุกต์มาจากโมเดลของสเปียร์แมน โดยได้ตั้งข้อสังเกตว่า คะแนนจริงของผู้สอบกับคะแนนที่วัดได้มีความสัมพันธ์ใกล้ชิดกันเพียงใด เพื่อเป็นการบอกจึงมีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนจริงกับคะแนนที่วัดได้ หรือเรียกว่า ดัชนีความเที่ยง นั่นเอง

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2$$

การประมาณค่าตวามเที่ยงนั้นสามารถกระทำได้หลายวิธี

**1. ความเที่ยงแบบคงที่** มักจะใช้วิธีการสอบซ้ำ แล้วประมาณค่าโดยคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนน

**2. ความเที่ยงแบบสมมูล** มักจะวัดเวลาเดียวกันโดยใช้แบบทดสอบแบบสมมูลแล้วประมาณค่าโดยคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนน

**3. ความเที่ยงแบบคงที่และสมมูล** มักจะใช้วิธีการวัดเวลาเดียวกันแบบสอบซ้ำโดยใช้แบบ ทดสอบแบบสมมูลค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนน

**4. ความเที่ยงแบบสอดคล้องภายใน** จะใช้กันอยู่หลายวิธีด้วยกัน ได้แก่ วิธีแบ่งครึ่งข้อสอบ วิธีคูเดอร์-ริชาร์ดสัน วิธีสัมประสิทธิ์แอลฟาของครอนบาค และวิธีวิเคราะห์ความแปรปรวนของฮอยท์ เป็นต้น

แต่ว่าความเที่ยงนั้นยังมีปัจจัยหลายประการที่ส่งผงต่อสัมประสิทธิ์ความเที่ยงนี้ ปัจจัยที่มีอิธผลและส่งผลสัมประสิทธิ์ความเที่ยง ได้แก่ ความเป็นเอกพันธ์ของกลุ่มผู้สอบ (Group Homogeneity) ความยาวของแบบทดสอบ (Test Length) ความสัมพันธ์ระหว่างข้อสอบ (Inter item Correlation) เวลาที่ในการทำข้อสอบ (Time Limit) และวิธีที่ใช้การประมาณค่าความเที่ยง (Methods of Estimating Reliabilities)

ปัจจัยเหล่านี้ควรนำมาพิจารณาเมื่อทำการศึกษาความเที่ยงของแบบสอบ โดยทั่วไปแล้วค่าสัมประสิทธิ์สหาสัมพันธ์จะสูงที่สุดเท่าที่จะเป็นไปได้ของแบบทดสอบนั้น

ความคลาดเคลื่อนมาตรฐานของการวัด (SEM) เป็นตัวบ่งชี้ถึงระดับการกระจายของคะแนนที่สังเกตได้จากคะแนนจริง การประมาณค่าความจริง สามารถกระทำได้โดยการประมาณเป็นช่วงด้วยการใช้ความคลาดเคลื่อนมาตรฐาน และประมาณเป็นจุดด้วยการประยุกต์การวิเคราะห์การถดถอย

**ความตรง (Validity)**

ความตรงเป็นคุณสมบัติที่สำคัญที่สุดของแบบสอบหรือของเครื่องมือวัดผล ซึ่งจะเกี่ยวข้องกับคุณภาพด้านความถูกต้องของผลที่ได้จากการวัด ทำให้เราสามารถนำคะแนนที่ได้ไปแปลความหมายถึงสิ่งที่มุ่งวัดได้อย่างเหมาะสม การตรวจสอบความตรง เป็น *“กระบวนการรวบรวมและวิเคราะห์หลักฐานเพื่อการสนับสนุนความเหมาะสมและความถูกต้องของการนำคะแนนจากเครื่องมือวัดไปสรุปอ้างอิงถึงคุณลักษณะที่มุ่งวัด”* การตรวจสอบความตรงสามารถจำแนกประเภทตามเป้าหมายสำคัญได้ 3 ประเภทหลักๆ ซึ่งได้แก่

1) การตรวจสอบความตรงตามเนื้อเรื่อง (Content validation)

2) การตรวจสอบความตรงตามเกณฑ์สัมพันธ์ (Criterion-related validation) และ

3) การตรวจสอบความตรงเชิงทฤษฎี (Construct validation)

**การตรวจสอบความตรงตามเนื้อเรื่อง (Content validation)** วิธีการตรวจสอบมักจะให้ผู้เชี่ยวชาญตรวจสอบความเหมาะสมของนิยาม และขอบเขตของเนื้อเรื่องหรือประสบการณ์ที่มุ่งวัด โดยหากต้องการตรวจสอบความสอดคล้องของข้อสอบกับจุดมุ่งหมายของการวัด มักจะใช้ IOC (Item- Objective Congruence) หากเป็นต้องการตรวจ สอบความสอดคล้องของข้อสอบกับเนื้อเรื่องที่มุ่งวัดมักจะใช้ CVR (Content Validity Ratio)

**การตรวจสอบความตรงตามเกณฑ์สัมพันธ์ (Criterion-related validation)**

ความตรงตามเกณฑ์สัมพันธ์เป็นคุณสมบัติด้านความสอดคล้องสัมพันธ์กันระหว่างคะแนนจากแบบสอบกับเกณฑ์ภายนอก (Criterion) ที่สามารถใช้วัดลักษณะที่ต้องการนั้นได้ ความตรงตามเกณฑ์สัมพันธ์สามารถจำแนกได้เป็น 2 ประเภทตามลักษณะของเกณฑ์ที่นำมาใช้หาความสัมพันธ์ของข้อสอบ ดังนี้

**1. ความตรงตามสภาพหรือความตรงร่วมสมัย (Concurrent Validity)** เป็นความตรงตามเกณฑ์ซึ่งสัมพันธ์กับสมรรถนะการดพเนินงานในสภาพปัจจุบัน โดยวิธีการประมาณค่าความตรงตามสภาพนั้นเราสามารถหาได้โดยการคำนวณจากค่าสัมประสิทธิ์สหาสัมพันธ์ระหว่างคะแนนจากแบบทดสอบกับเกณฑ์

**2. ความตรงเชิงทำนาย (Predictive Validity)** เป็นความตรงตามเกณฑ์ที่สัมพันธ์กับสมรรถนะการดำเนินงานในอนาคต การประมาณค่าความตรงเชิงทำนายของแบบทดสอบเราสามารถหาได้จากการคำนวณค่าสัมประสิทธิ์สัมพันธ์ระหว่าง คะแนนจากแบบทดสอบกับคะแนนเกณฑ์

จะเห็นได้ว่าความตรงเชิงทำนายกับความตรงตามสภาพ มีความแตกต่างกันที่ช่วงเวลาของการเก็บข้อมูลคะแนนเกณฑ์ ถ้าแบบทดสอบสามารถให้คะแนนสอดคล้องกับสภาพปัจจุบัน ถือว่ามีความตรงตามสภาพ แต่ถ้าสามารถให้คะแนนสอดคล้องกับสภาพที่เกิดขึ้นในอนาคตก็จะเป็นความตรงเชิงทำนาย ดังนั้นการมีความตรงเชิงทำนายย่อมมีความตรงเชิงสภาพ แต่ถ้าแบบสอบมีความตรงตามสภาพแล้ว ไม่จำเป็นต้องมีความตรงเชิงทำนายเสมอไป การประมาณค่าความตรงตามเกณฑ์สัมพันธีนั้นยังมีปัจจัยหลายปัจจัยที่ส่งผลต่อและเป็นข้อจำกัด เช่น ปัญหาเกี่ยวกับการเลือกเกณฑ์ ความเที่ยงขแงคะแยยสอบและคะแนนเกณฑ์และช่วงจำกัดของคะแนนสอบและคะแนนเกณฑ์

**การตรวจสอบความตรงเชิงทฤษฎี (Construct validation)** การตรวจสอบด้วยวิธีนี้จะแตกต่างจาก 2 วิธีแรก คือ การตรวจสอบความตรงตามเนื้อเรื่องและการตรวจสอบความตรงตามเกณฑ์สัมพันธ์ มุ่งที่จะทรายถึงความถูกต้องและเหมาะสมในการนำคะแนนสอบไปใช้ตามจุดมุ่งหมาย แต่ในขณะเดียวกันความตรงทั้ง 2 ประเภทนี้ต่างมีข้อจำกัดที่สำคัญหลายประการดังกล่าวมาแล้ว

ในกรณีที่ต้องการแปลงคะแนนเพื่อสรุปอ้างอิงลักษณะทั่วไปของบุคคล ซึ่งอาจเป็นลักษณะทางจิตวิทยา คุณลักษณะเหล่านี้ถือว่าเป็นโครงสร้างทางความคิด เป็นภาวะสันนิษฐาน ซึ่งเป็นนามธรรม ไม่สามารถสังเกตได้โดยตรง แบบทดสอบเหล่านี้จึงมีความจำเป็นต้องมีความตรงเชิงทฤษฎี

ความตรงเชิงทฤษฎี จัดว่าเป็นคุณสมบัติที่สำคัญที่สุดของเครื่องมือวัดลักษณะที่เป็นนามธรรม ที่ต้องทำการวัดทางอ้อม จึงจำเป็นต้องใช้การพิจารณาลักษณะในบริบทของทฤษฎี โดยอาศัย แนวคิดเชิงทฤษฎี (Theoretical concepts) สำหรับการนิยามลักษณะที่มุ่งวัด เสนอโครงสร้างการวัด แลพกำหนดแนวทางการตั้งสมมติฐาน

การตรวจสอบความตรงเชิงทฤษฎี เป็นกระบวนการศึกษาขอบเขตความเหมาะสมของการนำคะแนนจากแบบทดสอบไปแปลผลถึงลักษณะทางจิตวิทยาที่มุ่งวัดได้ดีเพียงใด ดังนั้นการตรวจสอบความตรงเชิงทฤษฎียังคงไม่สามารถยืนยันและสรุปได้จากการศึกษาผลการทำนายเพียงครั้งเดียว แต่ต้องการผลการตรวจสอบในแง่ต่างๆ ตามโครงสร้างทฤษฎีของลักษณะที่มุ่งวัดนั้นจนได้หลักฐานเชิงประจักษ์ที่ชี้ชัดและเชื่อถือได้อย่างคงเส้นคงวา

วิธีการตรวจสอบความตรงเชิงทฤษฎี จำเป็นต้องอาศัยหลักฐานความสอดคล้องกับทฤษฎี สำหรับวิธีการหาหลักฐานเชิงประจักษ์เพื่อสนับสนุนความตรงเชิงทฤษฎีหลายวิธี วิธีที่นิยมใช้กันมาก ได้แก่ วิธีตัดสินโดยผู้เชี่ยวชาญ วิธีเปรียบเทียบคะแนนระหว่างกลุ่มที่ทราบ วิธีการเปรียบเทียบคะแนนจากการทดลอง วิธีการวิเคราะห์เมทริกซ์พหุลักษณะ-พหุวิธี วิธีวิเคราะห์ตัวประกอบ

**การประยุกต์ใช้ Confirmatory Factor Model: CFM สำหรับตรวจสอบความตรงตามเนื้อเรื่องและความเท่าเทียมกันของเนื้อเรื่อง**

เนื่องด้วยการตรวจสอบความตรงตามเนื้อเรื่องมักประสบปัญหาหลากหลายอย่าง ในการสร้างแบบสอบคู่ขนาน จำเป็นที่แบบสอบเหล่านั้นจะต้องมีความตรงจามเนื้อเรื่อง และความเท่าเทียมกันของเนื้อเรื่อง ซึ่งสามารถประยุกต์โมเดล CFM มาช่วยในการตรวจสอบได้ตามขั้นตอนดังนี้

1) ตรวจสอบความตรงตามโครงสร้างของเนื้อเรื่องที่มุ่งวัดของแบบทดสอบแต่ละฉบับ

2) ตรวจสอบความเท่าเทียมกันของเนื้อเรื่องระหว่างแบบสอบ

**ก่อนเกิดเป็นทฤษฎีการทดสอบแนวใหม่**

ทฤษฎีการทดสอบแบบดั้งเดิม Classical Test Theory (CTT) ถูกใช้เป็นแนวคิดพื้นฐานสำหรับการพัฒนาแบบทดสอบทางการศึกษาและจิตวิทยาเป็นส่วนใหญ่ ตลอดช่วงศตวรรษที่ 20 ที่ผ่านมา ในปัจจุบันก็ยังมีการใช้กันอย่างกว้างขวาง แต่ด้วยข้อจำกัดหลายประการของการวัดตามทฤษฎีการทดสอบแบบดั้งเดิม ทำให้มีนักทฤษฎีทางการทดสอบหลายท่านได้ก่อกระแสการวัดคุณลักษณะภายในแนวใหม่ เพื่อที่จะแก้ไขจุดอ่อนเกี่ยวกับข้อตกลงเบื้องต้นของทฤษฎีการทดสอบแบบดั้งเดิม เริ่มต้นด้วยเทอร์สโตน (L.L. Thurston, 1927, 1928) เป็นผู้บุกเบิกการวัดคุณลักษณะภายในบุคคล และพัฒนาเทคนิคการวิเคราะห์ตัวประกอบสำหรับศึกษาคุณลักษณะทางจิตวิทยา อันเป็นการวางรากฐานความคิดที่สำคัญเกี่ยวกับ ทฤษฎีการทดสอบแนวใหม่ (Modern Test Theory: MTT) จากผลงานของ ครอนบาคและคณะ (Cronbach et.al., 1963, 1972) ได้เขียนบทความเรื่อง “The Dependability of Behavioral Measurement: Theory of Generalizability of Score and Profiles” ได้เสนอแนวคิดเกี่ยวกับโมเดลความเที่ยงทั่วไปภายใต้เงื่อนไขแบบต่างๆของการทดสอบ รวมทั้งผลงานของ ลอร์ดและโนวิค (Load and Novice, 1968) เรื่อง “Statistical Theories of Mental Test Score” ซึ่งได้นำเสนอหลักการและแนวคิดสำคัญที่ช่วยผลักดันกระแสการปฏิรูประบบความคิดของการวัดสู่ทฤษฎีการทดสอบแนวใหม่ รวมทั้งการมีส่วนร่วมของนักทฤษฎีการทดสอบในยุคใหม่อีกหลายท่าน จนได้พัฒนาการทดสอบแนวใหม่ซึ่งสามารถจำแนกออกเป็น 2 แนวทางสำคัญๆ ได้แก่ ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของการวัด (Generalizability Theory) หรือ G-Theory และทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) หรือ IRT แต่ละแนวทางมีแนวความคิดและความเป็นมาที่สำคัญดังนี้

**ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (Generalizability Theory)**

ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (Generalizability Theory) หรือ G-Study ได้มีการเสนอโมเดลของการศึกษาวิเคราะห์ความน่าเชื่อถือของผลการวัด หรือความเที่ยงในสถานการณ์หรือเงื่อนไขต่างๆ ของการวัด ทำให้เราทราบและสามารถควบคุมแหล่งความคลาดเคลื่อนได้ตรงประเด็น เพื่อผลการวัดมีความน่าเชื่อถือ หรือมีความเที่ยงสูงถึงระดับที่ต้องการ

จากแนวคิดทฤษฎีการทดสอบแบบดั้งเดิมมีความคิดพื้นฐานที่เชื่อว่า ความผันแปรของคะแนนที่สังเกตได้ประกอบด้วย 2 ส่วน คือ ความผันแปรของคะแนนจริง และความผันแปรของคะแนนความคลาดเคลื่อน ซึ่งไม่สามารถระบุหรือแบ่งแยกออกได้ นอกจากนี้ยังไม่ได้สนใจต่อเงื่อนไขการวัดและปฏิสัมพันธ์ระหว่างเงื่อนไขของการวัดที่สามรถส่งผลต่อความน่าเชื่อถือของการวัด ในขณะที่ G-Theory ได้เสนอแนวคิดในการแยกส่วนความคลาดเคลื่อน (Error) จากหลายแหล่ง (Multiple error sources) ประกอบด้วยความคลาดเคลื่อนอย่างเป็นระบบ (Systematic source) และความคลาดเคลื่อนแบบสุ่ม (Random source)

**แนวความคิดและข้อตกลงเบื้องต้น**

G-Theory เป็นทฤษฎีทางสถิติของการวิเคราะห์ความน่าเชื่อถือของผลการวัดในสถานการณ์ของการวัดผลลักษณะต่างๆ ที่เป็นเป้าหมายของการนำเครื่องมือไปใช้ ความถูกต้องความน่าเชื่อถือของการวัด คือ ความถูกต้องของการสรุปอ้างอิง (Generalization) จากคะแนนที่สังเกตได้ไปยังคะแนนจริงของบุคคล

การศึกษาความน่าเชื่อถือของผลการวัดตาม G-Theory ตั้งอยู่บนพื้นฐานของข้อตกลงเบื้องต้นดังนี้

1. คุณลักษณะที่มุ่งวัดของบุคคล เป็นค่าที่อยู่ในสภาวะคงที่ (Steady state)

2. ผู้สอบคนเดียวกันได้คะแนนต่างกันจากการวัดในแต่ละสถานการณ์ หรือเงื่อนไขของการวัดเนื่องมาจากความคลาดเคลื่อนที่เป็นระบบอย่างน้อย 1 แหล่ง

3. เมื่อพิจารณาผู้สอบทั้งกลุ่ม ความแปรปรวนที่สังเกตได้ ประกอบด้วย ความแปรปรวนของคะแนนจริง ซึ่งเป็นความแตกต่างที่แท้จริงระหว่างบุคคล

ผลการวิเคราะห์ด้วย G-Theory จะให้ค่าสัมประสิทธิ์สัมพันธ์เชิงสรุปที่แสดงถึงระดับความน่าเชื่อถือของคะแนนที่ได้จากการวัด (Level of dependability) เรียกว่า สัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของการวัด (Generalizability Coefficient) ซึ่งคล้ายกับสัมประสิทธิ์ความเที่ยง (Reliability Coefficient) ในทฤษฎีแบบดั้งเดิม

การตรวจสอบความน่าเชื่อถือของผลการวัดโดยใช้ G-Theory ทำให้ผู้บริหารการทดสอบสามารถทำการตัดสินความน่าเชื่อถือของผลการวัดได้ 2 ลักษณะ ได้แก่

1. ความน่าเชื่อถือของคะแนนในการนำผลไปใช้เชิงสัมพัทธ์ (Relative)

2. ความน่าเชื่อถือของคะแนนในการนำผลไปใช้สัมบูรณ์(Absolute)

**สัมประสิทธิ์การสรุปอ้างอิง (G-Coefficient)**

สำหรับทฤษฎี G-Theory สัมประสิทธิ์การสรุปอ้างอิง (G-Coefficient or $\rho_i^2$) เป็นสัดส่วนระหว่างความแปรปรวนของคะแนนเอกภพกับความแปรปรวนของค่าคาดหมายของคะแนนที่สังเกตได้ สัมประสิทธิ์การสรุปอ้างอิงมี 2 ประเภท คือ

1. สัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ ($\rho^2_{Abs}$)

2. สัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์ ($\rho^2_{Rel}$)

**การประมาณค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการออกแบบการวัดที่มีฟาเซทเดียว (G-Coefficient for Single-Facet Design)**

ในการศึกษาคุณภาพของแบบทดสอบ นักวัดผลต้องการทราบว่าสถานการณ์ของการใช้แบบสอบ หรือเงื่อนไขทดสอบ 1 ปัจจัย เช่น ความยาวข้อสอบ (I) หรือ จำนวนครั้ง (O) หรือจำนวนผู้ตรวจ (R) จะมีผลต่อความน่าเชื่อถือของแบบสอบเพียงใด

POPULATION: Examinees or Person

UNIVERSE: All possible combinations of conditions of the facet.

FACET: Item or Occations

DESIGN: p x i design

**การประมาณค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการออกแบบการวัดที่มีสองฟาเซท (G-Coefficient for Two Facet Design)**

**1.) การออกแบบ TWO FACET CROSSED DESIGN**

ในการศึกษาคุณภาพของแบบทดสอบ นักวัดผลต้องการทราบว่าสถานการณ์ของการใช้แบบสอบ หรือเงื่อนไขทดสอบ 2 ปัจจัย ได้แก่ ความยาวข้อสอบ (I) และ จำนวนครั้ง (O) จะมีผลต่อความน่าเชื่อถือของแบบสอบเพียงใด

POPULATION: Examinees or Person

UNIVERSE: All possible combinations of conditions of the facet.

FACET 1: Item (I)

FACET 2: Occations (O)

DESIGN: p x i x o design

**2.) การออกแบบ TWO FACET NESTED DESIGN**

ในการศึกษาคุณภาพของแบบทดสอบ นักวัดผลต้องการทราบว่าสถานการณ์ของการใช้แบบสอบ หรือเงื่อนไขทดสอบ 2 ปัจจัย ได้แก่ ความยาวข้อสอบ (I) และ จำนวนครั้ง (O) ของการทดสอบ (เมื่อการสอบแต่ละครั้งใช้ข้อสอบต่างกัน) จะมีผลต่อความน่าเชื่อถือของแบบสอบเพียงใด

POPULATION: Examinees or Person

UNIVERSE: FACET 1: Item (I)

FACET 2: Occations (O) และ i : o

DESIGN: p x (i : o) design

**การประมาณค่าสัมประสิทธิ์สรุปอ้างอิงสำหรับการออกแบบการวัดที่มีสามฟาเซท (G-Coefficient for Three Facet Design)**

ในการศึกษาคุณภาพของแบบทดสอบ นักวัดผลต้องการทราบว่าสถานการณ์ของการใช้แบบสอบ หรือ

เงื่อนไขทดสอบ 1 ปัจจัย เช่น ผู้ประเมิน (R) จำนวนครั้ง (O) และแหล่งของผู้ประเมิน (R) จะมีผลต่อความน่าเชื่อถือของแบบสอบเพียงใด

**ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT)**

ทฤษฎีการตอบสนองข้อสอบ หรือ IRT สามารถจำแนกออกได้เป็น 2 ลักษณะ ได้แก่ ทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนน 2 ค่า (Dichotomous IRT) และทฤษฎีการตอบสนองข้อสอบแบบตรวจให้คะแนนมากกว่า 2 ค่า (Polytomous IRT)

**แนวคิดพื้นฐานของ IRT**

โมเดลการตอบสนองข้อสอบ (IRT Models) เป็นระบบความสัมพันธ์ระหว่างตัวแปรอิสระที่รวมกันสำหรับทำนายตัวแปรตาม สำหรับโมเดลการตอบสนองข้อสอบ ตัวแปรอิสระจะประกอบด้วย ตัวแปรแฝง คือ ความสามารถที่แท้จริงของผู้สอบ ($\theta$) และคุณลักษณะของผู้สอบ (B) sหรือค่าพารามิเตอร์ข้อสอบ (a, b, c) ส่วนตัวแปรอิสระเป็นตัวแปรที่สังเกตได้ คือ โอกาสการตอบข้อสอบได้ถูกต้อง

โมเดลการตอบสนองข้อสอบจะประกอบด้วยค่าพารามิเตอร์และค่าคงที่ดังนี้

1. พารามิเตอร์ของผู้สอบ

$\theta$ = ระดับความสามารถผู้สอบ โดยที่ค่า $\theta$ มีพิสัยอยู่ระหว่าง $-\infty$ ถึง $+\infty$

2. พารามิเตอร์ของข้อสอบ

$a_i$ = ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบข้อที่ i (Discrimination parameter) ในทางทฤษฎีแล้วมีค่าระหว่าง $-\infty$ ถึง $+\infty$ แต่ในทางปฏิบัตินิยมใช้ $a_i$ ระหว่าง +0.50 ถึง +2.50

$b_i$ = ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i (Difficulty parameter) ในทางทฤษฎีแล้วมีค่าระหว่าง $-\infty$ ถึง $+\infty$ แต่ในทางปฏิบัตินิยมใช้ $b_i$ ระหว่าง -2.50 ถึง +2.50

$c_i$ = ค่าพารามิเตอร์โอกาสในการเดาข้อสอบถูก (Guessing parameter) ในทางทฤษฎีแล้วมีค่าระหว่าง 0 ถึง 1 แต่ในทางปฏิบัตินิยมใช้ $c_i$ ไม่เกิน 0.30

3. ค่าคงที่

e = 2.71828

D = 1.70

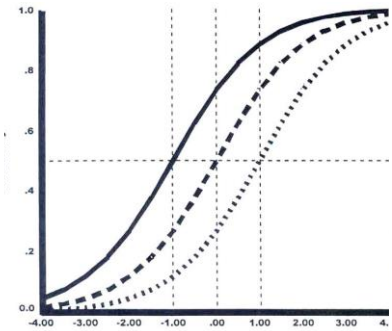
**ข้อตกลงของทฤษฎีการตอบข้อสอบ**

1. ความเป็นมิติเดียว (Unidimensionality: One trait)
2. ความเป็นอิสระ (Independence: Local Independent)
3. โค้งคุณลักษณะข้อสอบ (Item Characteristic Curve: Item Response Model)
4. ข้อสอบที่ใช้ต้องไม่เป็นข้อสอบประเภทความเร็ว

**โมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ (One parameter model or Rasch Model)**

โมเดลแรกที่มีความสำคัญที่สุดที่นำมาใช้ในโมเดลทฤษฎีการตอบข้อสอบ โค้งลักษณะข้อสอบสำหรับโมเดลโลจิสติกแบบ 1 พารามิเตอร์ แสดงในรูปฟังก์ชันทางคณิตศาสตร์ ดังนี้

$$P_i = \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}} \; ; \; I = 1,2,\ldots,n$$



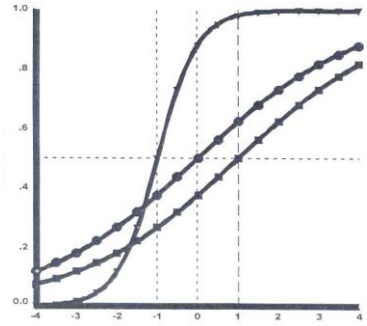

**รูปที่ 1** ลักษณะข้อสอบ (ICC) สำหรับโมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์

โค้งลักษณะข้อสอบ (ICC) สำหรับโมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ มีลักษณะที่สำคัญ คือ $b_i$ มีค่าแปรเปลี่ยนตามลักษณะข้อสอบแต่ละข้อ $a_i$ มีค่าคงที่ และ $c_i$ เท่ากับ 0

**โมเดลการตอบสนองข้อสอบแบบ 2 พารามิเตอร์ (Two parameter model)**

โมเดลแรกที่มีความสำคัญที่สุดที่นำมาใช้ในโมเดลทฤษฎีการตอบข้อสอบ โค้งลักษณะข้อสอบสำหรับโมเดลโลจิสติกแบบ 2 พารามิเตอร์ แสดงในรูปฟังก์ชันทางคณิตศาสตร์ ดังนี้

$$P_i = \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}} \text{ ; I =1,2,...,n}$$



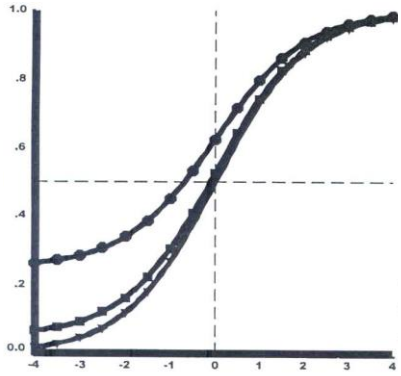

**รูปที่ 2** ลักษณะข้อสอบ (ICC) สำหรับโมเดลการตอบสนองข้อสอบแบบ 2 พารามิเตอร์

โค้งลักษณะข้อสอบ (ICC) สำหรับโมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ มีลักษณะที่สำคัญ คือ $b_i$ เป็นค่าที่แสดงตำแหน่งของ ICC ณ จุด $\theta$ ที่มีโอกาสตอบข้อสอบถูก 0.50 ส่วน $a_i$ เป็นค่าความชันของ ICC ณ จุด $b_i$ ส่วน $c_i$ เท่ากับ 0 และ e = 2.718, D = 1.7

**โมเดลการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ (Three parameter model)**

โมเดลแรกที่มีความสำคัญที่สุดที่นำมาใช้ในโมเดลทฤษฎีการตอบข้อสอบ โค้งลักษณะข้อสอบสำหรับโมเดลโลจิสติกแบบ 3 พารามิเตอร์ แสดงในรูปฟังก์ชันทางคณิตศาสตร์ ดังนี้

$$P_i = \boldsymbol{C_i} + (1-c_i)\frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}} \text{ ; I =1,2,...,n}$$

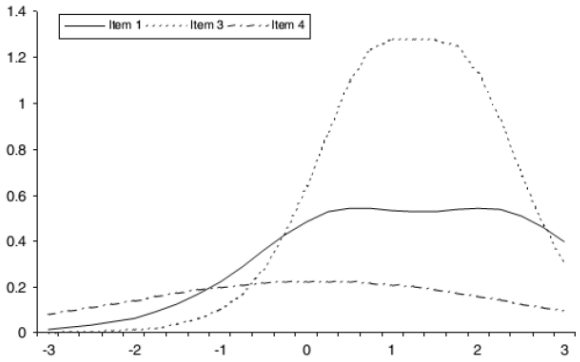



**รูปที่ 3** ลักษณะข้อสอบ (ICC) สำหรับโมเดลการตอบสนองข้อสอบแบบ 3 พารามิเตอร์

โค้งลักษณะข้อสอบ (ICC) สำหรับโมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ มีลักษณะที่สำคัญ คือ $b_i$ เป็นค่าที่แสดงตำแหน่งของ ICC ณ จุด $\theta$ ที่มีโอกาสตอบข้อสอบถูก $\frac{1+c_i}{2}$ ส่วน $a_i$ เป็นค่าความชันของ ICC ณ จุด $b_i$ ส่วน Ci คือค่าพารามิเตอร์โอกาสการเดาข้อสอบได้ถูก e = 2.718, D = 1.7

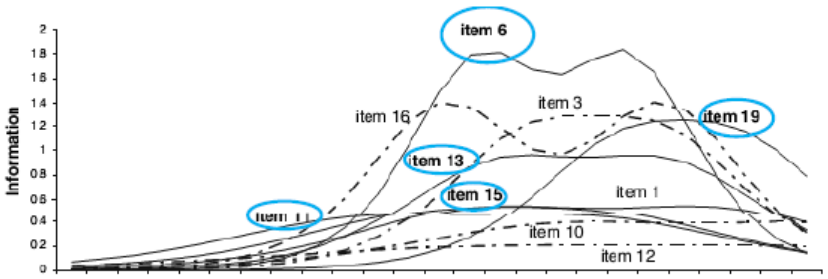
**ฟังก์ชันสารสนเทศของข้อสอบและแบบสอบ (Item and Test Information)**

**ฟังก์ชันสารสนเทศของข้อสอบ (Item Information)** เป็นดัชนีผสมที่สร้างจากัชนีคุณลักษณะของข้อสอบหลายลักษณะ ประกอบด้วย ค่าพารามิเตอร์อำนาจจำแนก ค่าพารามิเตอร์ความยาก ค่าความแปรปรวนคะแนนรายข้อ เพื่อใช้บ่งชี้คุณภาพของข้อสอบ




**รูปที่ 4** ลักษณะกราฟฟังก์ชันสารสนเทศของข้อสอบ (Item Information)

**ฟังก์ชันสารสนเทศของแบบสอบ (Test Information)** เป็นค่าฟังก์ชันสารสนเทศของแบบสอบอันเกิดจากผลรวมเชิงพีชคณิตของฟังก์ชันสารสนเทศของข้อสอบแต่ละข้อเข้าด้วยกันทั้งฉบับ ณ ตำแหน่ง $\theta$ เดียวกัน




**รูปที่ 5** ลักษณะกราฟฟังก์ชันสารสนเทศของแบบสอบ (Test Information)

**ความคลาดเคลื่อนมาตรฐานของการประมาณค่า (Standard Error of Estimation)**

ความคลาดเคลื่อนมาตรฐานของการประมาณค่า [SE($\theta$)] เป็นส่วนเบี่ยงเบนมาตรฐานของการแจกแจงความน่าจะเป็น ของค่าประมาณความสามารถที่แท้จริง ($\theta$) ซึ่งเป็นค่าสัดส่วนผกผันกับความถูกต้องแม่นยำของการประมาณค่าความสามารถ หรือค่าฟังก์ชันสารสนเทศของแบบสอบ

**โมเดลการตอบสนองข้อสอบแบบหลายมิติ (Multidimensional Item Response Theory)**

แนวคิดของการพัฒนาโมเดลการตอบสนองข้อสอบแบบหลายมิติ หรือ MIRT เป็นโมเดลที่มีคุณลักษณะแฝงมากกว่า 1 องค์ประกอบหรือหลายมิติ Multidimensional Item Response Theory ได้ขยายให้รองรับ $\theta$ หลายองค์ประกอบ

โมเดลการตอบสนองข้อสอบแบบหลายมิติถือว่าคุณลักษณะแฝงของบุคคลมากกว่า 1 องค์ประกอบส่งผลต่อการตอบข้อสอบ ดังนั้นพารามิเตอร์ของผู้สอบจึงมีตั้งแต่ 2 พารามิเตอร์ขึ้นไป การพิจารณาถึงความสามารถหลายมิติของบุคคล น่าจะช่วยทำให้โมเดลมีความสอดคล้องกับข้อมูลได้ดียิ่งขึ้น

จากการศึกษาเอกสารงานวิจัยที่ได้กล่าวไว้ข้างต้นแล้ว การวัดทางการศึกษาและจิตวิทยาเป็นการวัดคุณลักษณะภายในของมนุษย์ ซึ่งไม่สามารถสังเกตได้โดยตรง แต่มีความสำคัญและจำเป็นต้องศึกษา เพราะการวัดคุณลักษณะภายในจะทำให้เข้าใจการเกิดพฤติกรรมภายนอกของมนุษย์ที่สามารถสังเกตได้โดยตรง อันจะนำไปสู่การทำนายควบคุม และพัฒนาพฤติกรรมมนุษย์ การวัดคุณลักษณะภายในของมนุษย์จำเป็นต้องอาศัย ทฤษฎีการทดสอบ เพื่อทำความเข้าใจคุณลักษณะของสิ่งที่มุ่งวัด โครงสร้างของการวัด และการพัฒนาเครื่องมือสำหรับทดสอบ

ทฤษฎีการทดสอบเป็นองค์ความรู้ที่มีนัยทั่วไปเกี่ยวกับการทดสอบ วิธีการแก้ปัญหา การทดสอบและพัฒนาเครื่องมือการทดสอบ ซึ่งจะช่วยให้นักวัดผลสามารถทำการสร้างและพัฒนาแบบสอบให้มีคุณภาพ สามารถแปลความหมายผลการวัดได้อย่างถูกต้องและสามารถนำสารสนเทศไปใช้สำหรับ การตัดสินใจได้อย่างเหมาะสม ประเภททฤษฎีการทดสอบสามารถแบ่งได้เป็น 2 ประเภทหลักๆ ได้แก่ ทฤษฎีการทดสอบแบบดั้งเดิม ซึ่งจะเน้นการตรวจสอบความสัมพันธ์ระหว่างคะแนนที่สังเกตได้กับคะแนนที่แท้จรง โดยการวิเคราะห์คุณภาพรวมของข้อสอบและแบบสอบ ส่วนทฤษฎีการทดสอบแนวใหม่ มุ่งเน้นการขยายแนวคิดของทฤษฎีการทดสอบแบบดั้งเดิม เพื่อให้ได้แนวทางการวัดที่น่าเชื่อถือยิ่งขึ้นซึ่งมีพัฒนาการที่สำคัญ 2 แนวทาง ได้แก่ ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (Generalizability Theory) และทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT)

ทฤษฎีการทดสอบแนวใหม่ (Modern Test Theory) จะมุ่งเน้นการขยายแนวคิดของทฤษฎีการทดสอบแบบดั้งเดิมให้มีขอบเขตที่กว้างขวาง และมีความน่าเชื่อถือมากยิ่งขึ้น โดยปรับข้อตกลงให้มีความสมเหตุสมผล ตามความจริงมากขึ้น เกี่ยวกับความคลาดเคลื่อนของการวัด ยอมให้มีความคลาดเคลื่อนจากการวัดนั้นมีความแตกต่างกันออกไปตามสถานการณ์การวัด เพื่อศึกษาความเที่ยงทั่วไป ภายใต้เงื่อนไขของการวัดตามแนวของ Generalizability Theory: G-Theory รวมทั้งศึกษาความคลาดเคลื่อนจากการวัดที่มีความแตกต่างกันไปตามลักษณะของข้อสอบและแบบสอบ ขึ้นกับระดับความสามารถของผู้สอบ และของข้อสอบและแบบสอบ ทั้งยังพยายามวัดคุณลักษณะภายในหรือความสามารถที่แท้จริงของบุคคลตามแนว Item Response Theory (IRT)

ทฤษฎีการทดสอบแนวใหม่จึงน่าจะให้ผลการวัดที่ชัดเจนตรงประเด็นมากขึ้น อันจะเป็นประโยชน์ต่อการพัฒนาข้อสอบและแบบสอบให้สามารถนำไปใช้ในสถานการณ์ต่างๆของการทดสอบได้อย่างมีประสิทธิภาพ สามารถระบุนัยทั่วไปของคะแนนจริง คุณของแบบสอบตามเงื่อนไขของการทดสอบ รวมทั้งการประเมินความสามารถที่แท้จริงของผู้สอบ และการบรรยายพฤติกรรมการตอบสนองข้อสอบของผู้สอบได้เป็นอย่างดี

## เอกสารอ้างอิง

ศิริชัย กาญจนวาสี. (2555). *ทฤษฎีการทดสอบแนวใหม่* (พิมพ์ครั้งที่ 4). กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

ศิริชัย กาญจนวาสี. (2556). *ทฤษฎีการทดสอบแบบดั้งเดิม* (พิมพ์ครั้งที่ 7). กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

สำนักทดสอบทางการศึกษา สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ. (สิงหาคม 2558). *ความหมายและความสำคัญของการประเมินภาคปฏิบัติ.* คู่มือการพัฒนาความสามารถในการสร้างเครื่องมือประเมินภาคปฏิบัติ (Performance Assessment).

วิเชียร เกตุสิงห์, "หลักการสร้างและวิเคราะห์ข้อสอบ",กรุงเทพฯ: กองวิจัยการศึกษา สานักงานคณะกรรมการศึกษาแห่งชาติ, 2515.

ฉัตรศิริปยะพิมลสิทธิ์. “การวิเคราะหองคประกอบ,” ในวารสารการวัดผลการศึกษา. สำนักทดสอบทางการศึกษาและจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ. 20 (58) : พฤษภาคม-สิงหาคม, 2541.

Briggs, D. C., & Wilson, M. (2003). An introduction to multidimensional measurement using Rasch models. Journal of applied measurement, 4(1), 87-100.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart, and Winston.

Allen, Mary J. and Yen, Wendy M. Introduction to Measurement Theory. California: Brooks/Cole Publishing Company, 1979.

Crocker, Linda and Algina, James. (1986). Introduction to Classical and Modern Test Theory. New York: CBS College Publishing.